THE HUMAN HISTONE GENES: ORGANIZATION AND IN VITRO TRANSCRIPTION

BY

LUIS FELIPE SIERRA TIRADO

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1983

To Robin

ACKNOWLEDGEMENTS

First of all, I would like to thank Gary and Janet Stein for their help throughout the development of this thesis. I also want to thank all the members of the Department of Biochemistry and Molecular Biology for their support and encouragement. Especially, Drs. Richard Boyce, William Holloman and Peter McGuire will be remembered for their help and support. This work would not have been possible without the support and the friendship I found in all the members of our research group. Special thanks go to Dr. Gary McMaster for teaching me the basics and to Drs. Farhad Marashi and Mark Plumb for encouragement and helpful discussions, as well as warm friendship. Linda Green's technical and human qualities are most appreciated.

Most important, I wish to thank my wife Robin for her love and caring. Her contributions to the completion of this work are countless.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate Council of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

THE HUMAN HISTONE GENES: ORGANIZATION AND IN VITRO TRANSCRIPTION

By

Luis Felipe Sierra Tirado

April 1983

Chairman: Gary S. Stein
Major Department: Biochemistry and Molecular Biology

A human genomic DNA library contained in λCh4A phage was screened for histone genes. Several clones were identified, from which seven were further characterized. From these, two groups of three clones each were found to contain overlapping DNA fragments, since they shared common restriction maps, as well as histone gene organization patterns. A third type of arrangement was found to be present in clone λHHG 39. The identity of the different histone genes was originally determined by hybridization to heterologous DNA probes, and later confirmed by direct DNA sequencing of selected genes. Other investigators in our laboratory have confirmed the identity of the different histone genes by using a variety of methods.

The structural analysis of these clones indicated that, in humans, the histone genes are clustered, but no tandem repeats are readily apparent. This arrangement agrees with the findings of other authors with respect to the organization of histone genes in other vertebrates, such as Xenopus, mouse, chicken and man. Furthermore, it was found, through

hybridization to different DNA probes, that the human histone genes are interspersed with other transcribed sequences, including several members of the Alu family of DNA sequences. All of the histone genes present in HHG phage have been subcloned into pBR 322.

The H4 gene present in cloue λHHG 41 appears to code for one of the major species of H4 mRNA found in HeLa cells, since it can, upon hybridization, protect this H4 mRNA species over its entire length from degradation by S_1 nuclease. This gene has been transcribed in vitro, using the whole HeLa cell extract described by Manley. Using this system, and a series of 5' deletion mutants constructed by exonuclease digestion of a subclone containing the H4 gene plus flanking regions, it was found that no sequences upstream from the TATA box are required for the in vitro transcription of this gene. However, the same in vitro transcription system shows that sequences located as far as 800 base pairs downstream from the 3' end of the gene are required for the production of a run-off transcript. Accurate initiation of trancription in vitro can proceed in the absence of 3' flanking sequences.

INTRODUCTION

Each living organism requires, at least in its germ cells, the presence of its whole complement of genetic information in order to survive and reproduce. However, as initially demonstrated by Gurdon et al. (1), the germ cells are not the only ones that contain the whole complement of genetic material, but most somatic cells are also totipotential. Several exceptions to this observation have been described, most notably in the case of mammalian erythrocytes, which completely lose their genetic material as they differentiate from the mast cells to the mature red blood cell (2). Other exceptions include the rearrangement of immunoglobulin genes during lymphocyte differentiation (3), and the amplification of ribosomal RNA genes in the early developing embryos from several different species (4).

If most cells of an organism contain the whole set of DNA sequences, it is likely that the total DNA content per cell will be directly proportional to the position of any given organism in the evolutionary tree, the more complex and sophisticated organisms having necessarily more DNA per cell than the simpler ones. When this hypothesis was tested, not only was the statement found to be true (5,6), but when the content of DNA per cell was plotted vs. millions of years of evolution, a logarithmically growing curve was found (5,6). Such a curve suggests that higher organisms not only must code for a larger number of differentiated proteins in order to achieve their complexity, but they also must generate

a vast amount of DNA sequences that do not code for enzymes or structural proteins, but are rather thought to be involved in regulatory processes (5,6).

While fully differentiated eukaryotic cells contain large amounts of DNA in their nuclei, constitutive expression of all the genes present in the genome would most likely produce a totifunctional, undifferentiated cell, unless post-transcriptional control was operative in all cases. Since this mechanism of control would require the cell to spend large amounts of energy for this sole purpose, it is unlikely that post-transcriptional control is the sole way of regulating gene expression.

From these considerations, it becomes obvious that not all of the information present in any given cell in the form of DNA can be expressed into a final product, be it RNA or protein, at any given time. This fact is especially clear in the case of higher eukaryotic organisms, which are composed of many different organs and tissues, each of them having a specific function in the adult animal; that is, each cell has a commitment to produce only a specific set of differentiated proteins at any given time. Such a set of proteins can be as small as in the red blood cell, which, after differentiation, produces mainly one protein, i.e., globin, the rest of the genetic material being first repressed, and then lost (2); or it can be as complex as in the hepatocytes, which have a key role in intermediary as well as terminal metabolism and have high concentrations of many enzyme activities. Moreover, regulation of gene expression occurs not only as a result of final differentiation, but also as a response to a specific stimulus, as in the case of hormone-stimulated cells (7,8).

Regulation of Gene Expression

Regulation of gene expression, understood as the ability of a cell to decide whether or not to produce a final gene product, can reside at many levels, which can be grossly divided into transcriptional (9-11), post-transcriptional (12-14), translational (15-17) and post-translational (18-20). In prokaryotes, the most usual regulatory events seem to occur, as often happens in biological processes, at the first step in this chain of events, that is, at the transcriptional level (21-24). In eukaryotes, on the other hand, more processing steps are usually involved in the production of the final gene product, and even though transcriptional control has been invoked as being of primary importance in many systems (25-27), regulation at subsequent levels has also been suggested as playing a role in many other instances (28-31). In some specialized cases, transcriptional control might be achieved by modifications of the genetic material itself, through rearrangement, as in the case of the immunoglobulin genes (3), or amplification, as in the case of ribosomal genes (4).

In eukaryotes the primary product of transcription is usually, though not always, in the form of a long transcript known as heterogeneous nuclear RNA (HnRNA). A series of processing events then takes place in the nucleus, whereby the large HnRNA molecules are cut into smaller pieces (32,33); also, AMP residues are added to the 3' end of the molecule, to give rise to the poly (A) tail present in most mature mRNAs. About 80% of the total HnRNA is comprised of poly (A) containing RNA molecules (34). Addition of proteins also occurs at this step, giving rise to nuclear ribonucleoprotein particles (33,35,36). Other nuclear events include the capping of the 5' end (37), addition of internal methyl groups via specific methyl transferases (38,39), and splicing (40,41).

The mRNA has then to travel from the nucleus to the cytoplasm, by crossing the nuclear envelope. This process does not seem to be controlled by diffusion, but rather, association of the pre mRNA with proteins seems to be required (35,42). There is some controversy with regard to the nature of these proteins, since most of the evidence seems to indicate that the proteins bound to the HnRNA are different from those found in polysomal messenger ribonucleoprotein particles (mRNP) (36). Once in the cytoplasm, the mRNA undergoes its last processing step, namely, shortening of its poly (A) tail (32,43,44). This process seems to occur both before and during translation. All of the steps just mentioned can conceivably be used by the cell as post-transcriptional regulatory steps (45-46).

From all of these processes, probably the most interesting are the capping and splicing. Neither one of these two events has been found to occur in prokaryotic systems. Capping is accomplished through a long series of events, involving several nuclear enzymes, including a guanylyl transferase that adds a GMP moiety (from GTP) to the 5' $ppp^{A}/_{GpNp}$- terminus of the pre-mRNA; this guanine residue is added in a "backwards" manner, so that the product contains three phosphodiester linkages, in the form of $Gppp^{A}/_{GpNp}$-(47). This reaction is followed by the action of one or more methyl transferases, that add methyl groups at the N_7 position of the terminal guanosine residue, and, in some cases, at the 2'-O position of the penultimate ribose and the N_6 position of the adenine residue (47). Depending on the number of methyl groups present, cap structures can be separated into 3 groups (Cap 0, Cap 1 and Cap 2) (47). The function of the 5' cap present in eukaryotic mRNA is not completely clear; however, several studies have shown that the absence of

the cap on a mRNA significantly lowers its half life, probably by rendering the RNA susceptible to 5' exonucleases (48,49). Furthermore, the cap structure seems to be involved in the translational capacity of the mRNA, both in vivo and in vitro (50), since addition of free cap structures (m^7GpppA) to a translation mix inhibits the translation of both capped and uncapped mRNAs (50). Addition of S-adenosyl homocysteine (SAH) to the same translation systems inhibits the translation of uncapped mRNAs, but does not affect the translation ability of capped mRNAs, a fact that suggests that uncapped mRNA requires capping before it is adequately used in the translation system (50).

The splicing of pre mRNA molecules is a process that was not fully recognized until a few years ago. Post transcriptional cleavage of the initial transcripts from ribosomal RNA and tRNA genes has been known for some time. However, splicing of pre mRNA molecules was first observed, through electron microscopy of R-loops, in 1977 for adenoviruses (51). Upon hybridization of mature mRNA with the DNA template from which the mRNA was originally transcribed, areas of homology were observed as double stranded DNA/RNA hybrids. Surprisingly, there were defined, reproducible regions in which the DNA template "looped-out", indicating a lack of homology between the DNA template and the RNA transcript. This result could be explained in two alternative ways: 1) The DNA could be read uninterruptedly throughout the length of a pre-mRNA, that could then be spliced at specific points, to give rise to mRNA-size molecules. 2) The RNA polymerase molecule could read only those regions of the DNA that are required in the final mRNA, by a process of looping-out the DNA template during transcription. Several studies have confirmed the first hypothesis (52,53), and DNA sequencing studies have shown the general occurrence of

splicing in most, but not all mRNA-coding genes, the exceptions known so far being histone genes (54) and some of the interferon genes (55).

Regardless of whether or not any of these processes is actually involved in the regulation of gene expression, the possibility has to be considered, since no evidence to the contrary is presently available.

At translational and post-translational levels, regulation is also possible at many different points, such as the stability of a given mRNA (56), the presence of specific amino acyl-tRNA synthetases, initiation factors, or even, in some cases, the presence or absence of specific prosthetic groups (56,57). Regulation of the phenotypic expression of a specific gene can also be executed post-translationally, through protease or pH-dependent cleavages of the peptide molecule (58), enzymatic addition of chemical groups, like phosphate (59), acetyl (60), methyl (61), ADP-ribosyl (62) or others, and more subtly by modification of the tertiary or quaternary structure of the protein (63,64).

Histone Gene Expression

Our laboratory has been involved for a long time in the study of the regulation of histone gene expression throughout the cell cycle of cultured human cells. The histone genes represent an excellent model for the study of the regulation of genes which are not the expression of a finally differentiated state, but rather are turned on and off in a time-dependent fashion. Experiments reported by Borun *et al.* and others (65-67) have shown that histone mRNA sequences are present in the polyribosomes of cells preferentially during the S phase of the cell cycle. The synthesis of histone proteins also appears to be primarily restricted to the S phase of the cell cycle in several different cell lines (68-71), although a basal level of histone protein synthesis has

been observed throughout the cell cycle (72,73). Furthermore, there has been at least one report in which equivalent levels of histone protein biosynthesis was observed throughout the cell cycle (74). The fact that steady state histone mRNA levels approximately parallel the rate of histone protein synthesis in vivo, together with the fact that histone mRNA biosynthesis in vivo slightly precedes the peak of accumulation of histone mRNA in HeLa cells (75), indicates that the regulation of histone mRNA abundance occurs at least in part at the transcriptional level.

As a group, histones are among the most evolutionarily conserved proteins, especially in the case of H4 histone, which is the most conserved protein described so far. Undoubtedly, this high degree of evolutionary conservation relates to the fundamental structural role histone proteins play in assembling eukaryotic DNA into nucleosomes (76,77). Nevertheless, histone protein variants have been found. Within any single species, several variant histone proteins can be separated on high resolution gel systems (73,78,79). Variant histone proteins have been observed for H1, H2A, H2B and H3, but not for H4 (73,78,79). Similarly, several different mRNAs coding for the same or similar protein variants have also been observed (75), even in the case of H4 mRNA from HeLa cells, where no variant proteins have been detected (80). Results obtained by Dr. Mark Plumb in our laboratory, using hybrid selection of in vivo synthesized, ^{3}H-labelled RNA, have indicated that the expression of the genes which encode for different mRNA subclasses observed for H2A, H2B, H3 and H4 mRNA appear to be under coordinate control; that is, the apparent rate of synthesis is under the same temporal control for all of these genes (75). Even though this might not be true for the expression of the genes for basal histone synthesis outside the S phase of the cell

cycle (73), the fact that histone proteins are produced in stoichiometric amounts during the S phase of the cell cycle suggests that histone mRNA biosynthesis might be under coordinate control.

This idea was originally strengthened by the finding that, in sea urchins, all of the early histone genes are arranged in tandemly repeated clusters, each cluster containing one copy of each histone gene (reviewed in 81, see next section). One possible way of attaining coordinate control of genes that are clustered is through the production of a monocistronic mRNA. Because in sea urchins, all of the genes in any given cluster are arranged in the same polarity, this possibility is theoretically acceptable. However, experimental data have completely ruled out the possibility that the histone genes are transcribed as a polycistronic mRNA (82).

One purpose of this project was to isolate and characterize genomic clones containing human histone genes, in order to gain insight into their molecular structure and organization, as well as using them as tools for studying the mechanisms of control and expression of human histone genes. From the previous discussion, it is clearly of interest to learn if in humans, histone genes are also arranged as tandem repeats, or if they are arranged in a more dispersed fashion.

Histone Gene Organization in Different Species

The high G+C content, coupled with the high copy number (83) of the histone genes of several different species of sea urchin, allowed for their early purification (84) and cloning (85-88). The histone genes are repeated several hundred times in the sea urchin genome (83), and repeated CsCl-actinomycin D gradient centrifugations allowed their isolation as a satellite band (84). Analysis of histone gene DNA isolated

in this way indicated that, in sea urchins, the histone genes are clustered and tandemly repeated. All five histone genes were found to be separated from one another by short A+T rich DNA regions (84). Sea urchin histone gene clones have shown that the tandem repeats are around 6 Kb in length (with slight variation between different species), and have confirmed the idea that each gene is separated from the adjacent genes by a rather short (no more than about 1 Kb) A+T rich spacer (86,88-90).

Furthermore, detailed characterization of sea urchin genomic histone DNA clones indicated that all the coding regions within each repeat are arranged with the same polarity (91,92). In other words, the RNA transcripts are all produced from the same strand of DNA. Analysis of these clones indicated that the DNA coding regions are colinear with the mRNA transcripts, thus indicating that histone genes, in general, do not seem to contain intervening sequences (93,94). There is at least one instance in which an intron-containing histone gene has been found (95); however, it is not clear if this particular histone gene is transcriptionally active.

Although the most salient feature of histone gene organization in sea urchins is the presence of homogeneous, tandem repeats, some microheterogeneity in the composition of these clusters has been described (88,96-98), as well as the presence of "orphons", or histone genes that have rearranged and are found in different parts of the genome, separated from their parental repeat (99).

The tandem repeat arrangement of histone genes in sea urchins has led to speculations as to what the advantages of such an arrangement might be, since similar types of organization have been conserved in several species of sea urchins through millions of years of evolution. Tandemly repeated

clusters can, theoretically at least, facilitate a mechanism of coordinate control of the genes (100), thus enabling the organism to maintain a certain stoichiometry of the gene products. This argument would not explain, however, how the early developing sea urchin maintains a ratio of histone H1 to core histones (H2A, H2B, H3 and H4) of 1:2, as has been found to be the case (101), while the ratio of genes in the repeats is 1:1.

The histone genes of the fruit fly Drosophila melanogaster have been studied for some time (102), and have also been found to be clustered and tandemly repeated, with a repeat size of approximately 4.8 Kb. However, not only does the gene order differ from that found in sea urchins, but the genes are not arranged in the same polarity as in sea urchins (102). The H4 and H2B genes are read from the strand opposite to that from which the H1, H3 and H2A genes are read. The Drosophila histone genes are repeated about 100 times per haploid genome (102), and, superimposed on the major arrangement of the genes into tandemly repeated clusters, the presence of "orphons" has also been described (99).

In yeast, the histone genes are repeated only twice per haploid genome (103). This low copy number might be related to the small size of the yeast genome (104), which suggests that their need for histone protein synthesis per unit of time during S phase might be much lower than in the rapidly developing sea urchins or Drosophila, or in higher eukaryotes (103). The organization of histone genes in yeast is strikingly different from that observed in either sea urchins or Drosophila melanogaster. In yeast, the genes coding for histones H2A and H2B are adjacent to each other, but they are divergently transcribed. Furthermore, the second set of H2A and H2B genes, which are also adjacent to each other, is not in

the vicinity of the first H2A+H2B pair, but rather, they are separated by at least 35-60 Kb of unrelated DNA sequences. The two H2B genes of yeast encode two different H2B proteins, which differ from each other by four amino acids (105). Finally, genes coding for yeast histones H3 and H4 have not been detected in close proximity to the genes coding for H2A and H2B histone proteins (103).

In the case of the newt Notophthalamus viridescens, histone genes are repeated 600-800 times per haploid genome (106), and this species is known for having an extremely high DNA value (about 45 pg per haploid genome) (107).

It is also known that newt oocytes, like sea urchin oocytes, store large quantities of histone mRNA (108), a fact that might explain the lower need for rapid transcription of these genes in early development. All five histone genes are arranged in the Notophthalamus viridescens DNA as homogeneous 9 Kb clusters; however, the clusters are not tandemly repeated as in sea urchins, but are actually separated from each other by up to 50 Kb or more of unrelated DNA (106,109).

Xenopus histone genes are repeated 20 to 50 times per haploid genome (110). Several clones have been isolated (111-112), and their analysis seems to indicate that in these organisms, the histone genes are also clustered, however, there is extensive sequence divergence in the spacer regions of the clusters, and the gene order has been found to vary from cluster to cluster. Genomic blot analysis has shown, however, that there is a major repeat of histone genes in Xenopus. This repeat might contain up to 30 copies of each one of the four core histone genes (H1 has not been tested). The remaining genes appear to be organized in a highly idiosyncratic fashion, as it varies from individual to individual (113).

Several histone gene genomic clones have been isolated from chickens (114-118), some of which contain H1, and some of which contain H5, the variant protein that partially replaces H1 in adult avian erythrocytes. The general picture emerging from all these clones is that, in chickens, the histone genes are again clustered, but no apparent tandem repeat has been observed. The histone genes are repeated approximately 10 times per haploid genome in chickens (119).

The same basic pattern of histone gene organization has been observed in the case of the mouse. Analysis of several clones (120,121) has indicated the presence of clusters of histone genes, but with no apparent repeat. The same has been found to be true for human histone genes isolated independently in three different laboratories (118,122,123).

A more detailed description of the organization of human histone genes will be presented in the Results and the Discussion sections of this dissertation.

Transcription Studies Using Cloned DNA

The advent of recombinant DNA technology (124,125) has allowed scientists to do detailed structural and functional studies concerning nucleic acid metabolism. We now have the capacity to study the structure and function of isolated genes, much in the same way enzymology has advanced in the last 30 or 40 years. Many researchers have tried to dissect the anatomy of different genes, through *in vitro* manipulation of DNA sequences (reviewed in 126), and much insight has been gained by functional (transcriptional) analysis of DNA that has thus been modified.

A number of transcription systems that are dependent on exogenously added DNA templates have been described, the most currently used ones

being: 1) microinjection of cloned eukaryotic genes into Xenopus eggs (127,128) or oocytes (129-134), 2) transformation of cloned DNA sequences into mammalian cells in culture (135-138), 3) use of SV40-derived vectors for transfection of competent cell lines (139-142) and 4) in vitro transcription systems composed of soluble cellular extracts (143,144).

DNA sequences injected into the germinal vesicle of Xenopus eggs or oocytes are transcribed accurately for periods of up to five days (129). Transcription in eggs has been found to be much less efficient, although just as accurate, as oocyte transcription of exogenous DNA templates (130). Several DNA molecules, including poly d(A·T) (131), herpes virus thymidine kinase genes (132), SV40 (129), Drosophila histone genes (129), as well as sea urchin histone genes (145), among others, have been shown to be accurately and efficiently transcribed and processed (when appropriate) after injection into Xenopus oocytes.

The in vitro manipulation of several of these genes, prior to their injection into Xenopus oocytes, has allowed the identification of DNA sequences required for accurate transcription by Xenopus RNA polymerase within the oocyte. It should be emphasized that purified RNA polymerase does not effect the accurate transcription of any of these cloned sequences; additional factors are required. An analysis of the sequences required for accurate initiation of transcription will be presented in a later chapter.

DNA-mediated gene transfer (transformation) has been used to assay for phenotypic expression of selectable marker genes. The most commonly used methods involve the use of tk^- mouse L cells, which are transformed with a herpes simplex virus thymidine kinase gene (hsv-tk). Transformants are selected by growing the cells in HAT medium and cell clones are then

analyzed (135-138). HAT medium contains hypoxanthine, aminopterine and thymidine. Only TK^+ cells are able to grow in this medium (136). Since the selection is made through a phenotypic change, from TK^- to TK^+, it is clear that the transforming thymidine kinase gene not only gets expressed into mature mRNA, but this mRNA is properly translated into a functional protein (138). Transformation of the cells with other, non-selectable genes can be achieved by co-transformation of the gene of interest in the presence of an unlinked hsv-tk gene (136,137), or by transformation with a plasmid containing the hsv-tk gene linked to another gene of interest (135). Genes used to transform cells in this way have been found to be integrated into the cellular, high molecular weight DNA (136); however, no unique chromosomal location is apparent (138). By using this technology, mouse cells have been stably transformed with rabbit β-globin genes (135-137), ϕX174 (136), pBR 322 (136) and many other viral and eukaryotic genes. In general, transformation with any of these genes has led to a high level of integration of the gene into the genome, concomitant with the efficient production of properly processed, mature mRNA.

An alternative method of transformation is provided by the SV40-derived cloning vehicles (139-142). In competent cells, SV40 can be used in a vegetative form (in monkey cells, which are permissive), or as an integration vector (in human cells, which are semi-permissive) (140). When used to transfect monkey cells, eukaryotic gene-containing SV40 genomes can reach levels of about 100,000 copies per cell (141), so that the recipient cell contains the equivalent of a gene present at high copy number in an active chromosome (141). Mouse (139,141) and rabbit (140) α- and β-globin genes have been inserted into the late region of the SV40

genome, replacing the late gene coding for VP1 (146,147). Monkey cells transfected with these recombinant molecules transcribe, process and translate globin mRNA, either under viral promoter control (139) or globin promoter control (141).

A variation to these protocols has been introduced by P. Mellon and coworkers (142). They have constructed a vector, pSV0d, derived from pBR 322, but lacking the so-called poison sequences, and containing, in addition, the SV40 origin of replication. This plasmid can be used to transform Escherichia coli the same way pBR 322 does. However, when used to transform COS cells, a line derived from monkey CV-1 cells, which contains a constitutively expressed SV40 T antigen gene (148,149), this plasmid will replicate due to the binding of T antigen to the SV40 origin of replication. This system again produces transformed cells containing an active gene in a high copy number (142). No co-transforming gene, and no SV40 promoters are required for the expression of the gene of interest.

Recently, in vitro transcription systems that use cloned DNA as templates in the presence of soluble cell extracts have been described. A system developed in Roeder's laboratory utilizes an S-100 extract from a variety of cells (143), and its transcription of a cloned DNA template is dependent on the concomitant addition of rather large amounts of crudely purified RNA polymerase II. Another system, described by Manley and coworkers (144), utilizes a whole cell extract (150), that has been depleted of cellular DNA. This system does not require the addition of exogenous RNA polymerase II. Both of these systems have been successfully used to effect specific initiation of transcription in a wide variety of genes, including the adenovirus late (143,144,151,152) and early (152)

genes, conalbumin (141,152), ovalbumin (152), sea urchin histone H2A (145) and β-globin from mouse (153), rabbit (154) and humans (155).

Other methods of studying the transcription of cloned sequences have been described but have only been used in a limited number of cases. These include microinjection into somatic cell lines (156), liposome fusion (157,158) and erythrocyte fusion (159).

The use of the whole cell extract described by Manley et al. (144) to assess specific initiation of transcription of a human histone H4 gene will be described in this dissertation.

Transcription of Cloned Genes by RNA Polymerase III

Eukaryotic RNA polymerase III has been associated with the transcription of 5S RNA as well as tRNA genes (160) and some viral genes, such as the adenovirus VA gene (161).

The transcription of the 5S RNA genes from Xenopus is one of the best understood transcriptional processes to date. Both in Xenopus laevis and in Xenopus borealis, there are two types of 5S RNA genes. Oocyte-type genes are expressed only in oocytes, while somatic-type genes are expressed in most cell types, including the oocyte, where they are responsible for a low percentage of the total 5S transcripts (162). Both types of gene are reiterated in the Xenopus genome: there are about 20,000 copies of oocyte-type, and about 400 copies of somatic-type 5S genes per haploid genome (162). Both types of genes from both species of Xenopus have been cloned (163-165), and the genes appear to be clustered and separated from each other by about 80 nucleotides of A+T rich spacer DNA sequences (165). The final 5S RNA product appears to be identical with the primary transcript obtained after microinjection of Xenopus oocytes with cloned 5S genes (130,166).

Structural analysis of cloned 5S DNA sequences has indicated the conserved presence of the oligonucleotides AAAAG, AGAAG and GAC at 15, 25 and 35 nucleotides upstream from the transcription initiation site (165). The 10 bp spacing corresponds to one turn of the DNA double helix, a fact that suggests that these sequences might be involved in some DNA-protein interaction (165). However, deletion analysis performed both at the 5' end and the 3' end of a Xenopus borealis somatic gene (167,168) have indicated that accurate transcription of this gene can occur in the absence of these conserved flanking regions. Furthermore, these same studies have shown that only an intragenic DNA region is required for the transcription of this gene in an oocyte nuclear extract. The 5' end of this intragenic region is located between nucleotides +50 and +55, while the 3' end of the region is between nucleotides +80 and +83 (167,168).

These results are in agreement with those obtained by Engelke and co-workers (169), who used a foot-printing method (170) to determine that a purified factor extracted from Xenopus ovaries (TF III A) interacts with an intragenic region, covering from nucleotides 45 to 96, of the cloned 5S genes. This factor, a 37,000 D polypeptide, is necessary for oocyte and somatic 5S gene transcription but is not required for the transcription of a $tRNA^{Met}$ gene.

In the absence of this factor, purified RNA polymerase III fails to transcribe accurately the 5S genes (171-173). More recently, Gottesfeld and Bloomer (174) have shown that both the naked plasmids and plasmids reconstituted into chromatin in the presence of Xenopus oocyte extracts are transcribed efficiently and accurately in vitro. However, if the DNA is only reconstituted with purified histones, no in vitro transcription was observed. Furthermore, chromatin reconstituted only in the presence

of histones and TF III A was active as a template for in vitro transcription by purified RNA polymerase III (174).

The alanyl $tRNA_2$ gene from the silkworm Bombyx mori is also transcribed by RNA polymerase III (160). Microinjection of cloned genes into Xenopus oocytes (175), as well as in vitro transcription in heterologous (176) or homologous (177) extracts, has shown that, in this case, the primary transcript differs from the mature tRNA, and some processing is required. The primary transcript is 98 nucleotides long (175), and is processed into mature alanyl $tRNA_2$ by removal of the 5' triphosphate (175) and the first three nuclotides (176), removal of the terminal 22 nucleotides in a single, non-processing endonucleolytic step (175,176), and addition of a CCA motif at the newly generated 3' end (175).

As has been observed for 5S genes (178) and for several bacterial genes (179), termination of transcription of the Bombyx mori alanyl $tRNA_2$ gene occurs at a cluster of thymidines (176). Deletion studies have shown that transcription of the gene still occurs at detectable levels when all but 6 bp of 5' flanking regions have been deleted from the DNA template (175), and competition experiments have again suggested the interaction of a cellular component and an intragenic region of this template. In this case, however, a second DNA segment, located further than 11 bp upstream from the 5' end of the gene, seems to be involved in the efficient and accurate transcription of this gene by RNA polymerase III (177).

Transcription of Cloned Genes by RNA Polymerase II

RNA polymerase II is thought to be involved in the transcription of most or all eukaryotic mRNA molecules (160). I have already discussed the various systems currently available for the study of the sequences that

are required for the accurate transcription of cloned eukaryotic genes, both in vivo and in vitro. I will now discuss the information that these studies have provided, regarding putative in vivo and in vitro promoter sequences, as well as splicing, polyadenylation and termination signals.

The sequences required to promote transcription of cloned genes have been found to vary widely, according to the methodology used to assay for transcription. In general, in vivo studies have shown that the TATA box, located about 30 nucleotides upstream from the initiation site (180,181), is required for efficient transcription of a sea urchin H2A gene injected into Xenopus oocytes (145) or an α-globin gene transcribed in the pSV0d/COS cell system (142). The role played by the TATA box in in vitro transcription studies is less clear. It appears to be required for the in vitro transcription of a cloned conalbumin gene in a cytoplasmic S-100 extract (152), as well as a human β-globin gene transcribed in a whole HeLa cell extract (155). However, the TATA box has been found not to be essential for the in vitro transcription of a rabbit β-globin gene transcribed in a whole HeLa cell extract (154). Deletion of the TATA box preceding a sea urchin H2A gene reduced the efficiency of in vitro transcription by a factor of five; however, transcription was not abolished altogether (182). The same conclusion has been reached after deletion of the TATA box of SV40 early genes (183) and the polyoma virus early genes (184). In addition, point mutations in the TATA box of a sea urchin H2A gene (182) or a conalbumin gene (185) produce a marked decrease in the level of in vitro transcription of either gene. However, in neither case was transcription abolished altogether (182,185).

Both in vivo and in vitro studies have shown that deletion of the TATA box gives rise to the production of a heterogeneous population of transcripts. Analysis of these transcripts has shown that their 5' ends

are different, implying a role for the TATA box in directing the precise site of transcription initiation by RNA polymerase II, at a position about 30 nucleotides downstream from the TATA box, but independent of the nucleotide sequence at the actual cap site (145,154,181,182).

It is interesting to note that bacterial promoters also contain a sequence similar to the TATA box, at a position that precedes by 10 nucleotides the RNA start site (186). Deletion of these sequences completely abolishes transcription (187).

Other sequences upstream from the TATA box have been tentatively assigned promoter functions, based on their conservation between several related or unrelated genes. One of these sequences is the 5'-GGPyCAATCT-3', or "CAAT" box, described by Benoist et al. (188) and Efstratiadis et al. (181), and found between 70 and 80 nucleotides upstream from the start site of many genes. Deletion of the "CAAT" box does not decrease the in vitro transcription capacity of any gene tested so far, including a sea urchin H2A (145,182), conalbumin (152) and human or rabbit β-globin (154) genes.

Histone genes sequenced so far have shown a remarkable pattern with respect to the presence of "CAAT" boxes; H2A, H2B and H3 genes all contain "CAAT" boxes, although modified in some cases, while H1 and H4 genes lack these sequences (reviewed in 54). Evidence will be presented in this dissertation for the presence of at least one "CAAT" box at the 5' end of a human H4 histone gene.

Other sequences upstream from the "CAAT" box have been found to have an effect on the in vivo transcription of several genes (142,145,189,190); however, these sequences seem to have no effect on the in vitro transcription of most genes (154). An exception to this has been found in

the case of a sea urchin H2A gene, where deletion of a region comprising nucleotides -111 to -139 (starting from the cap site) seems to produce a down mutation when assayed *in vitro* (182).

There are also some sequence structures that appear to be characteristic of histone genes but have not been found in other RNA polymerase II-dependent genes. These include a 5'-GATCC-3' motif usually found about 10 bp upstream from the TATA box and a cap box (or mRNA initiation site) of the form 5'-PyCATTCPu-3' (191,192, reviewed in 54). On the other hand, the oligonucleotide 5'-CTTPyTG-3' often found slightly downstream from most cap sites (181,193,194) is not usually found in histone genes (reviewed in 54).

With regard to splicing of pre mRNA molecules into mature mRNA, little is known about the sequence requirements of the process. Many introns analyzed so far have been found to start at a GT at the 5' end and finish at an AG at their 3' end (195). Furthermore, point mutations introduced in these intron-exon junctions do seem to have an inhibitory effect on the cell's ability to accurately splice the mutated pre mRNA (196).

Most polyadenylated mRNA possess, close to their 3' ends, the sequence AAUAA, which is thought to be involved in the recognition by poly (A) polymerase or another factor involved in the polyadenylation of pre-mRNA molecules (197). The general model suggests that RNA polymerase II reads further downstream, through this polyadenylation site, and then, an endonuclease recognizes the poly (A) addition site and cleaves the molecule a few nucleotides downstream. This molecule will then serve as an appropriate *in vivo* substrate for poly (A) polymerase (181). Histone mRNAs derived from mammalian cells are usually not polyadenylated, and

lack the polyadenylation sequences just described (191,198, reviewed in 54). Exceptions are found in the yeast H4 gene, which produces a poly (A)-containing H4 mRNA, and it does contain the AAUAA motif about 40 nucleotides downstream from the stop codon (54), as well as egg histone mRNAs from a variety of species (17).

Histone genes, at variance with most other genes, seem to terminate transcription at the mature 3' end of the mRNA (54,199). This process seems to involve the recognition, either by RNA polymerase II or other factors, of a well conserved DNA sequence containing a hyphenated dyad symmetry, which is found close to the 3' end of histone mRNA (54). This sequence has been shown to be required for the termination of transcription of a sea urchin H2A gene in Xenopus oocytes (200). However, the same studies indicated that this sequence, although required for termination of transcription, is not sufficient, since insertion of the same sequence in the middle of a sea urchin H2B gene present in the same repeat did not produce premature termination of transcription of the H2B gene. This hyphenated dyad symmetry closely resembles prokaryotic promoter or attenuator sequences (129), as well as putative eukaryotic polymerase III terminators (165).

Histone genes contain a second homology block, a few nucleotides downstream from the ACCA termination motif, characterized by a high purine content (reviewed in 54). No specific function has been ascribed to these sequences.

More recently, a conserved region characteristic of the 5' end of histone H2B genes has been described, and again, no function for this highly conserved sequence has been determined (201).

The purpose of the work described in this dissertation was twofold: 1) Isolation and characterization of genomic clones containing human histone genes. 2) In vitro transcription of a human H4 gene. The first part of the work has allowed us a further understanding of the structure and organization of human histone genes (122,202). At the same time, it has provided our laboratory with a powerful tool with which to dissect the fundamental question of the levels and mechanisms of regulation of histone gene expression in human cell lines (68,75,203,204). The in vitro transcription system has been used to assess the sequence requirements for the accurate initiation of in vitro transcription of this H4 gene.

Results obtained during the first part of this work indicate that, in humans, histone genes are clustered, but not tandemly repeated (122). These genes also appear to be interspersed with other transcribed DNA sequences (202). In vitro transcription studies have indicated that no sequences upstream from the TATA box are required for accurate initiation, in accordance with what has been found in other systems. However, an unexpected involvement of regions downstream from the 3' end of the coding region in proper elongation of in vitro transcripts has been found.

MATERIALS AND METHODS

I. Isolation of Clones Containing Human Histone Genes

A. Growth of Phage and Bacteria:

λCh4A-derivative recombinant phage were grown in suspension in the DP_{50}.SupF strain of Escherichia coli by a modification of the PDS method described by F. Blattner et al. (205). A bacterial starter culture was grown overnight in 25 ml of NZYCM medium (1% casein hydrolysate or NZamine, 0.5% NaCl, 0.5% yeast extract, 0.1% casamino acids, 10 mM $MgSO_4$), supplemented with diaminopimelic acid (DAP) and thymidine (0.01% and 0.004%, respectively), as suggested by Maniatis' group (206,207).

Three milliliters from the overnight culture of DP_{50}.SupF bacteria were used to innoculate a 100 ml culture, and growth was followed by measuring A_{590}. When the optical density reached 0.5-0.6, 1 ml of bacteria was mixed with 1 ml of MgCa (10 mM $MgCl_2$/10 mM $CaCl_2$), and this mix was infected with an appropriate dilution of phage (usually 10^6-10^7 pfu). Incubation proceeded at 37°C for 5 min and the infected bacteria were diluted into 500 ml of pre-warmed NZYCM media, containing, in addition to DAP and thymidine, 1 ml of uninfected DP_{50}.Sup F bacteria. The culture was maintained at 37°C with vigorous agitation for 14 to 17 hours, or until lysis was apparent; 1 or 2 ml of chloroform were then added and the culture was stored at 4°C for a few minutes, before isolating the phage or spinning down the debris to prepare phage stocks.

Plasmid-bearing bacteria (E. coli strain HB 101) were grown in L-broth (1% casamino acids, 0.5% yeast extract, 0.5% NaCl, adjusted to pH 7-7.5 by addition of 2 ml of 1 M NaOH) containing 0.2% D-glucose, 10 mM $MgSO_4$ and 50 ug/ml ampicillin, and/or 25 ug/ml tetracycline. Five hundred milliliter of pre-warmed media were inoculated with 15 ml of an overnight starter culture that had reached stationary phase. Cells were grown at 37°C with moderate agitation for 3-4 hours, until they reached an A_{590} of 0.45-0.5. Plasmid amplification was then induced by addition of chloramphenicol to a concentration of 175 ug/ml (4.5 ml of a 20 mg/ml stock solution in 95% ethanol). The culture was allowed 16-18 hours of further incubation at 37°C, after which the cells were pelleted and plasmid DNA was isolated.

All experiments involving viable bacteriophage or bacteria containing recombinant DNA molecules were performed under conditions specified by the NIH Guidelines for Research Involving Recombinant DNA Molecules.

B. DNA Isolation

1. Phage DNA: Phage DNA was isolated by a modification of the method described by Blattner et al. (205). Infected cultures (usually 500 ml) were grown overnight until lysis was evident. One or two milliliters of chloroform were then added, followed by the addition of DNase I and RNase A to 100 ug/ml each. The lysates were then incubated at 37°C for 30 min and the bacterial debris was pelleted by two successive centrifugations at 7500 rpm for 30 min each in the Beckman JA-10 rotor. The

supernatant was adjusted to 0.5 M NaCl and 10% (w/v) PEG_{6000} (Eastman) and kept for at least 3 hours at 4°C with occasional stirring. Phage were then pelleted by centrifugation at 7500 rpm for 30 min. The pellet was carefully drained and then resuspended in 17 ml of ϕ80 buffer (205) containing 10 mM $MgCl_2$. Solid CsCl (8.5 gr) were added and the phage were then centrifuged through a pre-formed step gradient of CsCl. The gradient was prepared starting from the lightest solution, and carefully underlaying sequentially each of the higher density solutions. Accordingly, the phage solution was layered first, and then, using a pasteur pipet, 4 ml of a 1.45 gr/cc solution of CsCl were underlaid, followed by 4 ml of a 1.5 gr/cc solution, and finally, 7 ml of a 1.7 gr/cc solution of CsCl.

This gradient was centrifuged for 4 hours at 28,000 rpm in the Beckman Ti60 rotor. The phage band (observable by the naked eye approximately at the center of the tube) was extracted by puncturing the tube on the side, about 2 mm below the phage band, with a 16 ga needle attached to a 6 ml syringe. Usually about 2 or 3 ml of the gradient material were collected, and the phage were immediately layered on top of a pre-formed gradient containing 2 ml of each of the CsCl solutions used for the first gradient. The tubes were centrifuged overnight at 28,000 rpm in the Beckman Ti50 rotor. Phage were collected as before (usually 0.8-1 ml) and were extensively dialyzed against 10 mM Tris-HCl, pH 8.0/10 mM $MgCl_2$. At this point, the phage were disrupted by heating for 10 min at 68°C in the presence of 1% (w/v) sodium dodecyl sulfate (SDS). DNA was extracted twice with phenol equilibrated against 10 mM Tris-HCl, pH 8.0/1 mM EDTA (sodium ethylene diamino tetraacetic acid), and several times with chloroform:isoamyl alcohol ($CHCl_3$: IAA=24:1) (v/v). LiCl was added to a final concentration of 0.25 M and the DNA was precipitated overnight at

-20°C by addition of 2.5 volumes of cold ethanol. The ethanol precipitation was repeated once before resuspending the DNA at a concentration of 1 mg/ml in 10 mM Tris-HCl, pH 8.0/1 mM EDTA.

2. Plasmid DNA: Plasmid DNA was isolated by the cleared lysate-triton method (208). Bacterial pellets were washed once in 50 mM Tris-HCl, pH 8.0/ 25% (w/v) sucrose, centrifuged at 5000 rpm for 5 min in the Beckman JA-20 rotor and resuspended in the same buffer (10 ml per 500 ml of culture). After all the cells were in suspension, lysozyme was added to a final concentration of 200 ug/ml and the suspension was kept on ice for 5 min. Then EDTA was added to a concentration of 6.25 mM, followed by an equal volume of 50 mM Tris-HCl pH 8.0/6.25 mM EDTA/0.5% (v/v) triton X-100. The contents of the tube were carefully mixed and kept at 4°C for 20 min with periodic mixing. The mixture was heated for 10 min at 65°C and centrifuged for 30 min at 18000 rpm (31.000 X g) in the Beckman JA-20 rotor. An unstable pellet was obtained which contained bacterial DNA associated with denatured proteins. Plasmid DNA, RNA and soluble proteins remained in the supernatant, which was carefully drained into extraction tubes. The cleared solution was made 0.5% in SDS and extracted once with phenol, once with $CHCl_3$:IAA (24:1, v/v) and precipitated with ethanol in the presence of LiCl for 1-2 hours at -20°C.

The precipitated nucleic acids from one liter of culture were resuspended in 10 ml of 10 mM Tris-HCl, pH 8.0/1 mM EDTA, and ribonuclease A (previously made DNase-free by heating at 90°C for 10 min) was added to a concentration of 100 ug/ml. The solution was incubated at 37°C for 90 min with gentle agitation and then it was again extracted with phenol and with $CHCl_3$:IAA and precipitated twice with ethanol.

The DNA was then usually resuspended in 500 ul of 10 mM Tris-HCl, pH 8.0/ 1 mM EDTA and loaded on top of a BioGel A-15m chromatography column (30 X 1.5 cm). The chromatography was developed in the same buffer, and 1 ml fractions were collected. Elution of the DNA with the void volume of the column was followed by measuring the A_{260}. Fractions containing DNA (V_o, fractions 12-18) were pooled, ethanol precipitated in the presence of 0.25 M LiCl and resuspended at a concentration of 1 mg/ml in 10 mM Tris-HCl, pH 8.0/ 1 mM EDTA.

DNA obtained by this procedure consisted of a mixture of form I and form II plasmid DNA (usually in a ratio of 3:2), sometimes containing a slight contamination with bacterial DNA (never amounting to more than 1% of total DNA), as detected by electrophoretic analysis on 0.8% agarose gels.

In some cases, supercoiled DNA (Form I) was further purified by CsCl-ethidium bromide gradient centrifugation. In these cases, the DNA was not purified by BioGel A-15m chromatography, but rather, the DNA was diluted to 7.5 ml with 1 X SSC (1 X SSC: 0.15 M NaCl/0.015 M Na-citrate, pH 7.25) and mixed with 6.5 gr of solid CsCl and 100 ul of a 10 mg/ml solution of ethidium bromide. The tube was then filled with mineral oil and centrifuged for 48 hours at 40,000 rpm in the Beckman Ti60 rotor. DNA bands were visualized by transillumination with a long wavelength UV source and the band containing supercoiled DNA was extracted by puncturing the tube 2-3 mm below the band with a 16 ga needle attached to a 3 ml syringe. Ethidium bromide was immediately removed in the dark by passing the solution through a 2 X 0.8 cm Dowex (AG-150-X8) column, pre-neutralized with 1 X SSC. The column was washed with 1 ml of 1 X SSC and the eluate was dialyzed extensively against 10 mM Tris-HCl, pH 8.0/10 mM NaCl to remove the CsCl. DNA was then ethanol precipitated.

If pink coloration due to contamination with ethidium bromide persisted, one of two methods was used: a series of extractions with two volumes of n-butanol resulted in a reduction of the volume of aqueous solution, and was used when the pink color was detectable even before ethanol precipitation. If the coloration was weaker and could not be seen until the DNA had been concentrated by ethanol precipitation, then the pellet was resuspended in 300-500 ul of 10 mM Tris-HCl, pH 8.0/1 mM EDTA and then NaCl and EDTA were added to a concentration of 1 M and 0.1 M, respectively, bovine serum albumin (BSA) was added to a concentration of 100 ug/ml and the solution was thoroughly mixed before extraction with phenol and $CHCl_3$:IAA, followed by another ethanol precipitation.

C. Library Screening:

A human genomic DNA library was constructed by Dr. Tom Maniatis' laboratory in Caltech and kindly made available to us (207). In short, the library was constructed by partial digestion of human fetal liver DNA with restriction endonucleases Alu I and Hae III. These enzymes each recognize a 4 base pair sequence of DNA, and this digestion should produce a collection of quasi-random fragments of DNA. Fragments ranging between 15-20 Kb in length were isolated by sucrose gradient centrifugation. After protection of the internal Eco RI sites present in these molecules by treatment with Eco RI methylase, commercial Eco RI linkers were attached to the ends of the molecules by the use of T_4 DNA ligase, followed by complete digestion with Eco RI restriction endonuclease. The molecules were then ligated to Eco RI-digested λCh 4A arms (209), and the recombinant molecules were packaged *in vitro* and amplified in DP_{50}.Sup F bacteria (206,207). Lawn *et al.* (207) had estimated that a complete human genomic DNA library constructed this way should be contained in approximately 8×10^5 recombinant phage particles.

A complete equivalent of the human genomic DNA library (8 X 10^5 phage) was grown on solid agar plates and screened according to the technique described by Benton and Davis (210). Ten ml of NZYCM medium containing DAP, thymidine and 1.5% agar were poured into 15 cm diameter plastic petri dishes and allowed to solidify overnight. The next day, 1 ml from a stationary culture of DP_{50}.Sup F bacteria was mixed with 1 ml of MgCa and infected with 1 X 10^4 phage. After incubation at 37°C for 5 min, 8 ml of NZYCM medium containing DAP, thymidine and 0.7% agarose at 42°C were added and the tube was immediately inverted on top of an agar plate. The agar was allowed to solidify for 30 min at room temperature, and the plates were then incubated overnight at 37°C.

After incubation, the plates were allowed to cool at 4°C for 60 min. For filter lifting, the plates were removed from the cold box in groups of 10, nitrocellulose filters were carefully laid on top of the agarose and allowed to stay for 2 or 3 min. The orientation of the filters was established by puncturing 3 asymmetric holes through the filter and the agar with a needle containing india ink. Subsequently, the filters were removed and soaked for 20 seconds in 0.1 N NaOH/1.5 M NaCl, blotted on 3 MM paper and dipped for 20 seconds in 0.5 M Tris-HCl, pH 8.0/ 2 X SSC, blotted again and baked for 2 hours at 80°C in a vacuum oven.

After hybridization with a chicken genomic probe (vide infra) containing H3 and H4 histone genes, the agarose from areas of about 1 cm^2, corresponding to positive hybridization signals in the autoradiograms, was scraped out of the plate with a sterile pasteur pipet and the phage were allowed to diffuse into 1 ml of PSB (10 mM Tris-HCl, pH 7.4/100 mM NaCl/10 mM $MgCl_2$) (205) for 3 to 4 hours at 4°C. Then, appropriate dilutions were made in PSB (usually 10^{-3}-10^{-5}) and the

phage were plated in 9 cm petri dishes, exactly as described above for primary screenings. This process was repeated until phage obtained from a single positive plaque gave rise to a plate in which over 90% of the observable plaques showed positive signals upon hybridization.

D. Preparation of nick-translated probes:

1. Probes: The probe used for selecting human histone gene-containing recombinant λCh4A phage was a 2.6 Kb fragment containing chicken H3 and H4 genomic sequences inserted into pBR322, and kindly made available to us by Dr. Julian Wells (U. of Adelaide, South Australia). This DNA was nick-translated as explained in a later section.

2. DNA isolation from low gelling temperature agarose: The 2.6 Kb insert was separated from vector sequences by digestion with restriction endonuclease Hind III. Restriction endonucleases were purchased from BRL or from New England Biolabs, and were used as suggested by the supplier. The DNA fragments were then separated electrophoretically in a 0.8% low gelling temperature agarose gel and the DNA was isolated by the method of McMaster et al. (211). For this procedure, the agarose slice containing the DNA of interest was made into a paste with the help of a siliconized glass rod. NaCl was added to 0.5 M, EDTA to 10 mM and about 20 ug of yeast tRNA were added to serve as a carrier. The tube was then heated to 65°C for 10 min, vortexed for 5 seconds and incubated at 37°C for 5 min. Two volumes of phenol saturated with 0.5 M NaCl were added and the tube was withdrawn from the 37°C water bath, thoroughly mixed and centrifuged for 2-5 min in a microfuge at 4°C; the phenol extraction was repeated once, and the DNA was extracted once with $CHCl_3$:IAA (24:1) and ethanol precipitated in the presence of 0.25 M LiCl. The DNA was then resuspended, re-precipitated and finally resuspended again at an approximate concentration of 100 ug/ml.

3. Nick-translation: Probes were prepared by nick-translation. Nick-translation reactions were performed in Eppendorf tubes containing 80 uCi of dried [α-^{32}P] dCTP, in the presence of 50 mM Tris-HCl, pH 7.5/5 mM $MgCl_2$/1 mM β-mercapto ethanol/33 uM each of dATP, dGTP and dTTP. Usually 500 ng of DNA were nick-translated, and the reaction was started by addition of 6 units of DNA polymerase I from E. coli and 0.054 U/ml of DNase I. Incubations were for 60 min at 14°C, after which time the volume was increased to 200 ul by addition of 150 ul of 10 mM Tris-HCl, pH 8.0/1 mM EDTA.

The reaction mixture was then extracted once with $CHCl_3$:IAA (24:1) and the aqueous phase was loaded on a 9.5 X 0.9 cm BioGel A-15m column previously saturated with 100 ug of heat-denatured E. coli DNA. The chromatography was developed with 10 mM Tris-HCl, pH 8.0/1 mM EDTA, and the radioactivity present in the void volume of the column was determined by Cerenkov counting. Specific activities in the order of 10^8 cpm/ug were routinely obtained.

E. Hybridization:

Hybridizations to nitrocellulose-immobilized DNA were done essentially as described by Lawn et al. (207). Filters were washed for 10 to 15 min at room temperature in 4 X SET (1 X SET:0.15 M NaCl/2 mM EDTA/30 mM Tris-HCl, pH 8.0), and pre-hybridized for 60 min at 68°C in a volume ranging between 1.4 and 2 ml per 100 cm^2 of filter area in a solution containing 4 X SET/ 10 X Denhardt's/0.1% Na dodecyl sulfate (SDS)/0.1% Na pyrophosphate/100 ug/ml heat denatured E. coli DNA (1 X Denhardt's:0.02% polyvinyl pyrrolidone/0.02% ficoll/0.02% bovine serum albumin (BSA) (212)). The pre-hybridization solution was then replaced by a similar volume of the same solution containing, in addition, heat-denatured

radioactive probe, usually 1 X 10^6 cpm/ml. Hybridization was allowed to proceed at 68°C for 30-48 hours, after which time the probe was retrieved with a pasteur pipet.

The filters were washed three times for 20 min each at 68°C in 10 ml per 100 cm^2 of filter area with 4 X SET/0.1% SDS/0.1% Na pyrophosphate, then three times for 20 min each at 68°C in a similar volume of 1 X SET/0.1% SDS/ 0.1% Na pyrophosphate. Finally, the filters were washed once in a large volume of 2 X SET at room temperature, blotted between Whatman 3 MM paper and exposed while wet to Kodak XAR-5 or Cronex X-ray film at -70°C.

Filters were consecutively used for hybridizations with up to 4 or 5 different probes. For this purpose, probes were removed by dipping the wet filter in boiling water for 3 min, followed by a 10 sec wash in cold water. Appropriate elution of the old probe was monitored by exposing the filter to X-ray film overnight.

II. Characterization of Histone Gene-Containing Clones

A. Restriction Mapping:

Human DNA contained within several λCh 4A phage was mapped with respect to restriction endonucleases Eco RI, Hind III and Bam HI recognition sites, and selected subclones prepared by insertion of Eco RI fragments into pBR 322 (see below) were further mapped with respect to several other restriction endonucleases. In both cases, the approach consisted of digesting the DNA to completion with each one of the enzymes, both singly and in all possible combinations of two enzymes. The DNA fragments produced were electrophoretically fractionated on 0.8% agarose and/or 5% polyacrylamide gels, and the molecular weight of the fragments

were determined by plotting their migration against the migration of molecular weight markers obtained from Hind III-digested λDNA or from pBR 322 digested with Hinf I. In many cases, the order of overlapping fragments was facilitated by information gained by hybridizing Southern blots of the same gels to specific histone DNA probes (detailed below). In cases where results were unclear due to the presence of small molecular weight DNA fragments that stained poorly with ethidium bromide, visualization of restriction fragments was facilitated by labelling the DNA at the 3' end with the Klenow fragment of E. coli DNA polymerase (213). One hundred nanograms of DNA were labelled in a volume of 10 ul in the presence of 6 mM Tris, pH 7.5/6 mM $MgCl_2$/6 mM β-mercaptoethanol/50 mM NaCl/100 ug/ml BSA/ 33 uM each of dATP, dGTP and dTTP and 0.5 uCi [α-^{32}P] dCTP. The reaction was started by addition of 0.7 units of Klenow fragment of E. coli DNA polymerase, and incubation proceeded for 10 min at room temperature (about 22°C). Then 20 ug of yeast tRNA were added, together with 9 volumes of 0.3 M Na acetate. Nucleic acids were precipitated by the addition of 2.5 volumes of ethanol and, after 30 min in dry ice, they were pelleted by a 10 min centrifugation in a microfuge at 4°C. The pellet was washed once with 70% ethanol, radioactivity was determined by Cerenkov counting (214) and the DNA was analyzed in an appropriate gel.

B. Gel Electrophoresis of DNA:

Agarose gels of different concentrations (ranging from 0.5% to 3.0% agarose) were prepared and run in Tris-acetate buffer (40 mM Tris-HCl/5 mM Na acetate/ 1 mM EDTA, adjusted to pH 7.8 with glacial acetic acid). Horizontal beds were used in all cases, and the gels were usually electrophoresed at 125 mA (100 V) for 3 to 4 hours, although specific conditions varied widely, according to the purpose of the gel, as well as physical convenience.

Polyacrylamide gels were run in a vertical apparatus in TBE buffer (50 mM Tris-HCl/50 mM boric acid/ 1 mM EDTA, pH 8.3). After degassing the gel was polymerized by the simultaneous addition of TEMED (N,N,N',N' tetra methyl-ethylenediamine) and ammonium persulfate to concentrations of 0.0075 and 0.075% respectively. These gels were also run at 125 mA.

If the DNA was not radioactive, gels were stained for 10 min in 2.5 ug/ml of ethidium bromide, and DNA bands were visualized by exposure to long wavelength UV light. Photographic recording of the gels was obtained with a Polaroid Land Camera, using Type 57 Polaroid film.

C. Detection of Histone Coding Regions:

1. Southern blotting: Phage or plasmid DNA digested with appropriate restriction endonucleases was electrophoretically fractionated in 0.8% agarose gels as previously described. The gel was then stained with ethidium bromide, photographed and the DNA was transferred to nitrocellulose by the method of Southern (215). The gel was soaked for 20 min in 0.1 N NaOH/1.5 M NaCl, then for 30 min in 3 M NaCl/0.5 M Tris-HCl, pH 7.0. Meanwhile, a piece of nitrocellulose was cut to the same size of the gel and floated over 2 X SSC for 5-10 min. A thick sponge was saturated with 20 X SSC inside a deep plastic pan, and the gel was laid on a 3 MM paper on top of the sponge. The nitrocellulose filter was then carefully laid on top of the gel and covered with two layers of 3 MM paper and sufficient paper towels to absorb the 20 X SSC from the pan. Osmotic transfer was allowed to occur for 20-24 hours, with at least one change in the pad of paper towels. After transfer was complete the nitrocellulose filter was baked for 2 hours at 80°C in a vacuum oven. The extent of transfer was verified by staining the gel with ethidium bromide and observation under a long wavelength UV source.

2. cDNA: Among the first probes used to hybridize to Southern blots containing λHHG phage DNA, was a cDNA we prepared to 7-11 S polysomal RNA from S phase HeLa cells. Because histone mRNAs are among the main components of this fraction (65,216,217), hybridization with λHHG phage DNA would suggest the presence of histone genes among the λHHG phage. 7-11S RNA from S phase HeLa S_3 cells was polyadenylated using ATP-polynucleotidyl exotransferase from maize in a reaction containing 70 mM Tris-HCl, pH 8.8/1 mM ATP/10 mM dithiothreitol/1 mM $MnCl_2$. Polyadenylated RNA was reverse transcribed by AMV reverse transcriptase (kindly provided by Dr. J. Beard) in the presence of [α-^{32}P]dCTP in a reaction containing 40 ug/ml RNA/50 mM Tris-HCl, pH 8.3/20 uM β-mercaptoethanol/10 mM $MgCl_2$/30 mM NaCl/20 ug/ml oligo (dT)/ 50 ug/ml actinomycin D/1 mM each of dATP, dGTP and dTTP/30 uM dCTP, and 200 units/ml reverse transcriptase.

3. Heterologous DNA probes: Other probes used to characterize the λ HHG phage were DNA fragments obtained by digestion of different recombinant plasmids with appropriate restriction endonucleases, followed by isolation from low gelling temperature agarose gels as described, nick translation in the presence of [α-^{32}P]dCTP, and hybridization under the same conditions used to screen the library. Specific fragments used will be described in the Results section.

D. Subcloning into pBR 322:

1. Phage DNA digestion: Almost every Eco RI fragment derived from the inserts of each of the seven λHHG phage described has been subcloned into the Eco RI site of pBR 322. For this purpose, 3 ug of DNA from each λ HHG phage were digested with Eco RI. After confirming by gel electrophoresis that the digestion had been carried out to completion,

between 200 and 500 ng of λHHG DNA (200 ng of λHHG 6, λHHG 17, λHHG 22, λ HHG 41 and λHHG 55, 300 ng of λHHG 5 and 500 ng of λHHG 39) were ligated to 700 ng of Eco RI-digested, calf intestine alkaline phosphatase-treated pBR 322. This allowed a molar ratio of about 10:1 between the vector and the Eco RI fragments derived from each one of the phage.

2. Calf intestine alkaline phosphatase treatment of pBR 322: Two micrograms of pBR 322 DNA were digested to completion with an excess of Eco RI restriction endonuclease, phenol extracted once, $CHCl_3$:IAA (24:1) extracted once and ethanol precipitated in the presence of 0.25 M LiCl. The DNA was then resuspended in 20 ul of 10 mM Tris-HCl, pH 8.0/0.1 mM EDTA, and incubated at 65°C for 30 min with two units of calf intestine alkaline phosphatase. Two more units of enzyme were added, and incubation was continued for 30 min more at 65°C. Proteinase K was added to a final concentration of 1 mg/ml and the mixture was incubated for 30 min at 37°C. Then SDS was added to 0.25% and incubation was allowed for 15 min more at 37°C. The solution was then made 0.5% in SDS and extracted with phenol, $CHCl_3$:IAA (24:1) and precipitated twice with two volumes of ethanol in the presence of 0.25 M LiCl.

3. Ligation: DNA fragments to be ligated were mixed at the appropriate ratios (see above) in 66 mM Tris-HCl, pH 7.6/6.6 mM $MgCl_2$/10 mM dithiothreitol (DTT)/1 mM ATP. The reaction was started by the addition of two Weiss units of T_4 DNA ligase (218), and incubation proceeded at 12°C for 4 hours. Under these conditions, intermolecular ligation has been favored by the relatively high concentration of DNA molecules, and circularization was then favored by diluting the reaction mixture 10-fold in the same buffer. After incubating for 30 min on ice, five more units of T_4 DNA ligase were added, and incubation at 12°C was allowed to proceed overnight.

The ligation mixture was heated to 65°C for 5 min before transformation of E. coli cells, to disrupt reassociated but unligated molecules.

4. Transformation of E. coli strain HB 101: Twenty-five milliliters of L-broth were infected with 125 ul of an overnight culture of E. coli strain HB 101 and grown with agitation until the A_{590} reached exactly 0.45. Bacteria (2 ml) were taken in a sterile conical centrifuge tube, centrifuged at 4000 rpm for 5 min, carefully drained and resuspended in 500 ul of ice cold 30 mM $CaCl_2$ by gentle vortexing. Bacteria were then incubated in ice-water for 10 min and 100 ng of ligated DNA were added. Incubation in ice-water was continued for 30 min, after which the cells were heat-shocked at 43°C for 90 seconds. 2 ml of sterile media were added and the cells were allowed to recover at 37°C for 90 min with vigorous shaking before plating 200 ul aliquots on freshly made L broth-agar plates containing 50 ug/ml of ampicillin.

5. Selection: Colonies were allowed to grow overnight, after which they were transferred to duplicate plates by using a sterile toothpick. One of the plates contained 50 ug/ml of ampicillin, and the other contained 50 ug/ml of ampicillin plus 25 ug/ml of tetracycline. Colonies that were able to grow on the ampicillin plates, but not in the ampicillin plus tetracycline plates were further analyzed.

In some cases, the selection of colonies containing an appropriate insert was done by hybridization to nick-translated probes. For this, the colonies were transferred to a nitrocellulose filter (or, in some cases, grown on top of a nitrocellulose filter placed over an agar plate). The filter was then processed by a modification of the Grunstein-Hogness procedure (219). The filter was placed on Whatman 3 MM papers saturated with 0.5 N NaOH for 7 min, blotted, and applied to Whatman 3 MM papers

saturated with 1 M Tris-HCl, pH 7.4 for 1 min, followed by the same solution for 5 min. The filters were then blotted again, placed on Whatman 3 MM papers equilibrated with 1.5 M NaCl/0.5 M Tris-HCl, pH 7.4 for 5 min and finally shaken for 10 min in 2 X SSC.

After baking at 80°C under vacuum for 2 hours, the filters were washed in 4 X SET for 10 min at room temperature, prehybridized for 60 min at 68°C in 4 X SET/1 X Denhardt's/0.1% SDS/0.1% Na pyrophosphate and then boiled in double distilled water for 5 min to remove bacterial debris. The filters were then blotted on 3 MM paper, washed in 4 X SET for 10 min at room temperature, prehybridized, hybridized and washed as previously described.

E. DNA Sequencing:

1. Strategy: A pBR 322 plasmid containing a human genomic insert bearing an H4 gene was isolated, and the location of the H4 gene was determined by Aleida Leza to reside primarily within a 330 bp Sac II fragment. Preliminary in vitro transcription studies had shown that the 5' end of the mRNA was probably located within the adjacent Eco RI/Sac II fragment. To obtain the complete sequence of the mRNA plus its flanking regions, both of these fragments were sequenced starting at all four available 5'ends. This was accomplished by digesting DNA with both restriction endonucleases, followed by BAP,kinase and strand separation, as described in the next section.

2. ^{32}P-Labeling and Strand Separation: ^{32}P-labelling of Eco RI/Sac II-digested pFO 108A DNA was done by kinasing 40 ug of DNA by the protocol described under section III. E. 1. a. This amount of DNA represents approximately 6.7 pmoles of DNA, or 40 pmoles of ends.

After ^{32}P-labelling, the DNA was resuspended in 60 ul of 30% dimethyl sulfoxide/1 mM EDTA/0.05% xylene cyanol FF (XC)/0.05% Bromo Phenol Blue (BPB). Samples were denatured at 90°C for 2 min, followed by a quick chilling in dry ice. The samples were immediately loaded on a 4% polyacrylamide gel, containing a 50:1 ratio of acrylamide to bis-acrylamide. Electrophoresis was performed at 200 V for 12 hours (220).

The gel was exposed to Cronex X-ray film while wet, the desired bands were sliced out and the DNA was electroeluted by a modification of published procedures, as described in Section III.E.1.b. (221). The pellets were washed once with 300 ul of 70% ethanol, dried under vacuum and resuspended in 35 ul of water.

3. Sequencing Reactions and Sequencing Gels: DNA sequencing was done by the method of Maxam and Gilbert (220). The reactions used were those for specific cleavage at guanines (G), total purines (A+G), total pyrimidines (C+T), cytosines (C) and adenine > cytosine (A). Sequencing reactions were done exactly as suggested by Maxam and Gilbert and will not be described in detail.

Sequencing reaction products were analyzed on acrylamide-urea gels (220). For the first 30 or 40 nucleotides starting from the labelled 5'end, 20% gels were used, while for nucleotides 20 to 200, 8% gels were utilized. Both types of gels contained deionized urea at a concentration of 8.3 M. Urea was deionized by stirring for 10 min a 10 M solution of urea in the presence of 2 gr of activated charcoal and 2 gr of mixed bed resin (AG 501-X8, from Bio Rad Laboratories) per 100 ml of solution. The solution was cleared by filtration through a glass filter. Gels were 45 cm (20%) or 90 cm long (8%), and 0.3 mm thick. Electrophoresis was

performed at 2000 V inside a room kept at 37°C, in order to maintain a constant surface temperature of 52°C. The wet gels were then developed by autoradiography on Kodak XAR-5 or Cronex film for 3-7 days at -20°C.

III. In Vitro Transcription

A. Preparation of Whole Cell Lysate:

A whole cell extract was prepared from continuously dividing HeLa S_3 cells according to the protocol of Manley (144). Extracts prepared in this manner have been shown to initiate accurately, but not to terminate, transcription by RNA polymerase II from a variety of eukaryotic or viral promoters such as adenovirus (143,144,151,152), β-globin (153-155), ovalbumin (152), etc. The cells were harvested by centrifugation at 1500 rpm for 5 min at 37°C, washed twice in about 30 ml per liter of cell culture of ice cold Spinner salts (Gibco), followed each time by centrifugation at 1500 rpm for 5 min at 4°C. All the following steps in the extract preparation were carefully done at 4°C. The cells were resuspended in four times the packed volume of cells (which is about 1.8 ml per liter of cells at a density of 6-6.5 X 10^5 cells/ml) of 10 mM Tris-HCl, pH 7.9/1 mM EDTA/5 mM DTT. Cells were allowed to swell on ice for 20 min and then they were lysed by 8 strokes with a "B" pestle (tight) of a Dounce manual homogenizer. Lysis was followed by direct observation of the cells under a phase microscope. After lysis, four times the packed volume of cells of 50 mM Tris-HCl, pH 7.9/10 mM $MgCl_2$/2 mM DTT/25% sucrose/50% (v/v) glycerol were added, followed by the dropwise addition of one packed volume of cells of saturated, cold ammonium sulfate. The slurry was gently stirred on ice for 20 min, followed by a 3 hour centrifugation at 40000 rpm in the Beckman Ti 60 rotor to remove the

chromatin and precipitated proteins. The supernatant was then carefully removed and its volume measured. One gram of solid ammonium sulfate was added per 3 ml of supernatant; after the ammonium sulfate had dissolved, the solution was neutralized by addition of 1 ul of 1 M NaOH per gram of ammonium sulfate. The mix was then stirred for 30 min on ice. The second ammonium sulfate cut was collected by centrifugation at 15000 X g (11500 rpm in the Beckman JA-20 rotor), the precipitate was resuspended in one tenth of its volume of 20 mM HEPES, pH 7.9/100 mM KCl/12.5 mM $MgCl_2$/0.1 mM EDTA/2 mM DTT/ 17% (v/v) glycerol and then dialyzed twice for 4-8 hours against 100 volumes of the same buffer. Finally, the dialysate was centrifuged at 10,000 X g (9100 rpm in the Beckman JA-20) for 10 min and 200 ul aliquots were stored frozen in liquid nitrogen.

B. In vitro Transcription:

Several conditions for in vitro transcription have been reported, and it appears to be the concensus that different genes require slightly different conditions for optimal transcription in vitro. In particular, variables usually include lysate concentration (varying between 10 and 30 ul per 50 ul reaction), DNA concentration (between 10 and 75 ug/ml) and NTP concentration (between 300 uM and 1 mM). Glycerol, $MgCl_2$ and KCl concentrations have been varied by several investigators, and some recommend the use of phosphocreatine and/or an excess of ATP (144,152,155).

Several of these variables were tested, using as a template Eco RI-digested pFO 108 DNA, a clone containing a human H4 gene, as well as at least one member of the Alu I family of repetitive DNA sequences (202). The conditions found to be optimal for transcription of this clone were as

follows: 30 ul of lysate, 2.5 ug of DNA (50 ug/ml final), 1 ul of 7 mM EDTA (0.2 mM final), 1 ul of 50 mM phosphocreatine (1 mM final), 5 ul of 10 mM NTP (1 mM final of ATP, GTP and CTP, 0.05 mM UTP) and 20 uCi of [$-^{32}$P] UTP in a total volume of 50 ul. -amanitin was used at a concentration of 2 ug/ml when required. In some cases, reaction mixes were only 25 ul, and everything was reduced accordingly. Transcription was allowed to proceed for 50 min at 30°C. At this time, cold UTP was added to a concentration of 1 mM, and incubation was continued for 15 min at 30°C, to chase partially synthesized, labeled RNA molecules into full-size transcripts.

C. Isolation of in vitro Transcription Products:

In vitro transcription reactions (in either 25 or 50 ul) were terminated by addition of 55 ul of 10% SDS and 195 ul of fresh 2 mM Tris-HCl, pH 7.4/1 mM EDTA/2 ug/ml polyvinyl sulfate/1 ug/ul proteinase K. The contents of the tube were vortexed gently and digestion by the protease was allowed to proceed for 15 min at room temperature. The solution was then adjusted to 0.25 M NaCl by addition of 15 ul of 5 M NaCl, and nucleic acids were isolated by one extraction with phenol:$CHCl_3$:IAA (25:24:1), followed by one extraction with $CHCl_3$:IAA (24:1). After precipitation in dry ice for 15 min with 3 volumes of ethanol in the presence of 0.25 M LiCl, the precipitate was centrifuged for 15 min in a microfuge and resuspended in 150 ul of 0.2% SDS, followed by addition of 150 ul of 2 M ammonium acetate. Nucleic acids were precipitated again with 3 volumes of ethanol, and, after centrifugation, the precipitates were washed once with 70% ethanol. Incorporated counts were determined by direct Cerenkov counting (214).

D. Analysis of in vitro Transcripts on Formaldehyde Gels:

Samples were resuspended in 16 ul of 50% (v/v) formamide/6% (v/v) formaldehyde/50 mM Na_2SO_4/1 mM EDTA, heated at 70°C for 5 min and quick chilled. 8 ul of dye (50% (v/v) glycerol/50% (v/v) formamide/0.05% bromo phenol blue (BPB)/ 0.05% xylene cyanole FF (XC)) were then added, and the sample was applied to a pre-electrophoresed (30 min at 45 mA) 1.5% agarose/3% (v/v) formaldehyde gel. Electrophoresis was performed at 50 mA (about 65 V) for 3.5-4 hours in sample buffer minus formamide and containing only 3% formaldehyde (222).

The gel was then dried and exposed to XAR-5 X-ray film, usually overnight.

E. Detection of Specific Initiation of Transcription:

Specificity of initiation of transcription at the correct in vivo 5' end of the H4 mRNA by the in vitro transcription system was originally assayed by direct sizing of α-amanitin sensitive transcripts with different 3' end points (obtained by restriction enzyme digestion) in 1.5% agarose/3% (v/v) formaldehyde gels as described in the previous section.

A more accurate definition of the 5'end of the in vitro synthesized RNAs was attempted by a primer extension method (155). In short, this method is based on hybridization between the RNA to be analyzed and a DNA fragment, labelled at its 5'end and containing only sequences internal to the RNA. This DNA is then used as a primer by AMV reverse transcriptase, which transcribes the RNA into DNA until the 5'end of the RNA is reached. The elongated DNA fragment is then analyzed on a suitable gel (155).

1. Preparation of primer: The DNA used as primer was a 64 bp Alu I/Hha I fragment from pFO 108 A, containing sequences encoding from amino acid 17 (Arg) to amino acid 38 (Ala) of the H4 protein encoded in pFO 108 DNA.

a. BAP/kinase: Twenty micrograms pFO 108 A DNA were digested to completion with Alu I restriction endonuclease, followed by phenol extraction, $CHCl_3$:IAA (24:1) extraction and ethanol precipitation. The DNA was resuspended in 50 ul of 0.1 M NaCl and digested with 4.8 units of heat treated bacterial alkaline phosphatase for 15 min at 37°C, followed by 30 min at 60°C. EDTA was added to 1 mM and the sample was heated at 70°C for 5 min, phenol extracted, $CHCl_3$:IAA (24:1) extracted and dialyzed overnight against 2.5 mM Tris-HCl, pH 9.2. The dialysate was dried under vacuum to a volume of 45 ul. Five microliters of 10 X kinase buffer (0.5 M Tris-HCl, pH 9.2/0.1 M $MgCl_2$/ 0.05 M DTT/50% glycerol) and 1 ul of 0.1 M spermidine were added, and the sample was heated at 70°C for 5 min. The sample was then transferred to an Eppendorf tube containing 120 uCi of dry [γ-^{32}P]ATP, and 10 units of T_4 polynucleotide kinase were added. The reaction was allowed to proceed for 30 min at 37°C, after which 0.25M EDTA was added to 1 mM. The sample was again heated at 70°C for 10 min, and then it was phenol extracted once, $CHCl_3$:IAA (24:1) extracted once and ethanol precipitated.

b. Electroelution: The DNA was digested for 60 minutes at 37°C with an excess of Hha I restriction endonuclease (5 units per ug of DNA) and the 64 bp fragment was isolated from a 10% polyacrylamide gel by electroelution (221). The gel slice was placed inside a dialysis bag containing 3 ml of 0.25 X TBE buffer plus 20 ug of yeast tRNA. The dialysis bag was placed parallel to the electrodes of a horizontal electrophoresis apparatus containing 0.5 X TBE buffer. Electroelution was for 2.5 hours at 200 V, followed by 2 min at 200 V with the polarity reversed. The eluate was removed from the bag, extracted once with $CHCl_3$:IAA (24:1) and ethanol precipitated twice. Particulate matter was removed by filtration through siliconized glass wool and the DNA was precipitated again with ethanol.

2. RNA samples: RNAs transcribed in vitro using 10 ug of restriction endonuclease-digested pFO 108 A DNA as a template were isolated as previously described. After the last wash with 70% ethanol, samples were resuspended in 500 ul of 10 mM Tris-HCl, pH 7.5/2 mM $CaCl_2$/10 mM $MgCl_2$/2 ug/ul polyvinyl sufate, and heated at 100°C for 5 min. After addition of 250 ug of yeast tRNA as carrier, the DNA template was digested with 25 ug of RNase-free DNase I for 10 min at 37°C. The sample was made 10 mM in EDTA, phenol extracted, $CHCl_3$:IAA (24:1) extracted and ethanol precipitated in the presence of 0.25 M LiCl. RNase-free DNase I was prepared by pre-incubating a 1 mg/ml solution of DNase I in 20 mM Tris-HCl, pH 8.0/10 mM $CaCl_2$ at 37°C for 20 min, followed by addition of proteinase K in the same buffer to a final concentration of 1 mg/ml. Incubation was for 2 hours at 37°C, after which the DNase was used directly (223).

Polysomal RNA (7-11 S) from HeLa S_3 cells, provided by Dr. Farhad Marashi, was used as a control for primer extension experiments.

3. Hybridization and reverse transcription: DNA fragments, in vitro transcription products and control RNA were all resuspended 0.1 M NaCl. DNA and RNA were mixed in a total volume of 20 ul, denatured by heating at 100°C for 5 min, and transferred quickly to a 60°C water bath. Incubation was continued for 60 min, after which the water bath was turned off, thus allowing the samples to cool slowly to 40°C.

Reverse transcription was done in the same buffer as previously described, (Section II.C.2) except that [α-^{32}P]dCTP was omitted, and cold dCTP was added to a final concentration of 1 mM. The final volume was 50 ul. After starting the reactions by addition of 10 units of AMV reverse transcriptase, the reaction was allowed to proceed for 45 min at 37°C.

4. Analysis of primer extension products: The RNA used as a template for reverse transcription was degraded by incubation in 0.1 M EDTA/0.2% SDS/0.3 M NaOH for 60 min at 50°C (224). The solution was adjusted to 0.5 M Na acetate, phenol extracted, $CHCl_3$:IAA (24:1) extracted and ethanol precipitated. Samples were analyzed in 10% polyacrylamide/8.3 M urea gels run for 6 hrs at 17 W (27 mA). Urea was removed from the gel by two 15 min washes in 50% ethanol and the gel was then dried and exposed to X-ray film for autoradiography.

F. Construction of 5' Deletion Mutants from pFO 108 A:

1. BAL-31 exonuclease digestion: Ten micrograms of pFO 108A DNA were digested to completion with Eco RI restriction endonuclease, phenol extracted, $CHCl_3$:IAA (24:1) extracted and ethanol precipitated.

Nuclease BAL-31 from Alteromonas espejiana BAL-31 was purchased from BRL and used as suggested. The rate of removal of nucleotides from the free ends of duplex DNA was calculated according to the formula proposed by Gray et al (225). The actual rate under the conditions used in the experiment was confirmed by Klenow labelling (Section IIA) an aliquot from the reaction, followed by digestion with Hind III restriction endonuclease and analysis on 0.8% agarose gels. The theoretical and experimental values were in excellent agreement. The reaction was set in a volume of 400 ul, containing 20 mM Tris-HCl, pH 8.1/200 mM NaCl/12 mM $CaCl_2$/12 mM $MgCl_2$/1 mM EDTA. Ten micrograms of Eco RI-digested pFO 108A DNA were digested with 0.2 units of BAL-31 (0.5 units/ml) at 30°C for a total time of 15 min. Aliquots of 13 ul were withdrawn every 30 seconds, pooled in groups of 6 and frozen in the presence of 20 mM EDTA/20 mM EGTA. Samples were phenol extracted, $CHCl_3$:IAA (24:1) extracted and ethanol precipitated.

2. Cloning: Three hundred nanograms of BAL-31 digested DNA were resuspended in 60 ul of 60 mM Tris-HCl, pH 7.9/6 mM $MgCl_2$/20 mM

-mercaptoethanol/1 mM ATP/100 ug/ml BSA and ligated to Eco RI linkers labeled by kinasing in the presence of [γ-^{32}P]ATP at a ratio of linkers to DNA ends of 50:1. The reaction was started by addition of 3.3 Weiss units of T_4 DNA ligase. Ligation was performed at 14°C for 4 hours, after which time the NaCl concentration was adjusted to 50 mM, and the DNA was extensively digested with Eco RI and Hind III restriction endonucleases over a period of 5 hours, with periodic additions of Eco RI enzyme (a total of 200 units of Eco RI was used). DNA was then phenol extracted, $CHCl_3$:IAA (24:1) extracted and ethanol precipitated. The samples were resuspended in 50 ul of 10 mM Tris-HCl, pH 8.0/1 mM EDTA and the ligated DNA was separated from excess linkers by chromatography on a 9.5 X 0.9 cm BioGel A-15m column. Radioactive fractions eluting in the void volume were collected and ethanol precipitated in the presence of 20 ug/ml of yeast tRNA and 0.25 M LiCl.

DNA molecules were then ligated to EcoRI/Hind III-digested pBR 322 DNA, that had previously been treated with calf intestine alkaline phosphatase, as described in Section II.D.2. One hundred nanograms of the digested DNA were ligated to 500 ng of vector in a 100 ul reaction containing 66 mM Tris-HCl, pH 7.6/6.6 mM $MgCl_2$/10 mM dithiothreitol/1 mM ATP and 7.7 Weiss units of T_4 DNA ligase. The reaction was allowed to proceed overnight at 12°C.

Transformation of E. coli strain HB 101 was done exactly as described in Section II.D.4.

3. In vitro transcription: DNA purified from 5' deletion mutants was digested with Hind III or with Eco RI and Hind III restriction endonucleases, phenol extracted, $CHCl_3$:IAA (24:1) extracted, ethanol precipitated and transcribed in vitro as described in section III.B., both in the presence and absence of 4 ug/ml of α-amanitin.

RESULTS

Library Screening

Recent developments in molecular biology have indicated the advantage of using recombinant DNA technology to produce reliable probes for understanding gene expression and organization. With this concept in mind, we screened a human genomic DNA library, searching for histone genes. The human genomic DNA library was constructed by Dr. R.M. Lawn et al., and kindly made available to us by Dr. T. Maniatis. These investigators have calculated that a complete equivalent of the gene library should be contained in approximately 8×10^5 recombinant phage (207). At a mean of 20 Kb of human DNA per recombinant phage, that represents 1.6×10^{10} bp of human DNA, that is, five times more than the estimated size of the human genome (3×10^9) (226). Due to the quasi-random method of digestion of the human DNA used to construct the library, these reseachers postulated that screening one equivalent of the library, that is, 8×10^5 recombinant phage, should allow the isolation of any given sequence of human DNA. Although several investigators have found that screening one equivalent of any given library does not necessarily yield a clone containing the sequence being sought (118,227), it was nevertheless decided that, in searching for human histone genes, screening 8×10^5 phage should be enough, since the histone genes are repeated 20-40 fold in the haploid human genome (227).

Aliquots of DP_{50}.Sup F bacteria that had been previously infected with 1 X 10^4 phage were plated on 15 cm diameter petri dishes. After allowing for growth overnight at 37°C, the phage DNA was transferred to nitrocellulose filters and hybridized with ^{32}P-labelled DNA obtained from the insert present in p2.6H, a plasmid containing chicken genomic sequences coding for histones H3 and H4 (118). An example of the results obtained is shown in Figure 1A. This primary screening gave 109 plaques showing positive hybridization signals. Phage present in each one of these areas were isolated by impaling them from the plates with a Pasteur pipet, dissolved in PSB buffer and grown again, this time on 9 cm diameter petri dishes, for a second screening. Plates containing between 20-100 plaques were selected for transfer and hybridization. Twenty-eight of the original plaques showed clear positive signals in this secondary screening, while several others showed weak or unclear results, and were not pursued any further. For each of the 28 clear positives, plaques were selected from areas of the plates where a single positive plaque was found, with no negative plaques in close proximity. A third screening was then performed, in order to insure the purity of the isolated clones. In this third screening, an example of which is shown in Figure 1B, twenty-four clones showed positive hybridization signals for more than 90% of the observable plaques, indicating that the isolates were pure.

Clones were then grown in liquid, and the DNA was isolated, digested with Eco RI restriction endonuclease, and the fragments were electrophoretically fractionated on a 0.8% agarose gel. The DNA was transferred to nitrocellulose and hybridized to the same DNA fragment used to screen the library, containing chicken histone H3 and H4 genomic DNA sequences. Figure 2 shows the results obtained with fifteen different clones, all of which show positive hybridization, although very weak in

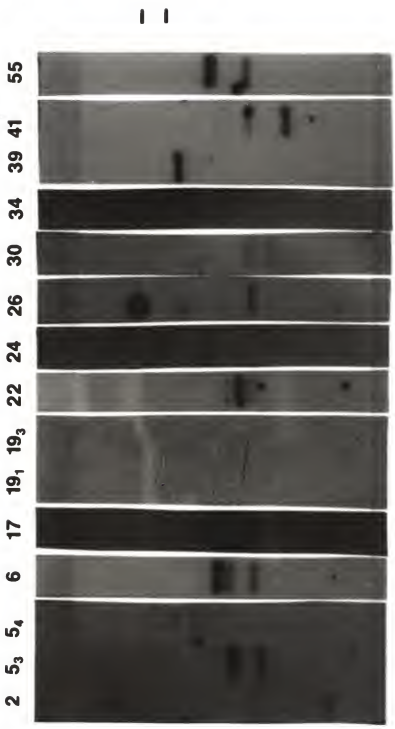
Figure 1. Screening of a human genomic DNA library.

A whole human genomic DNA library, cloned into λCH4A phage was screened by the Benton and Davis procedure (210). 10,000 phage were grown on each 15 cm diameter petri plate for the primary screening. A. A representative plate from the primary screening, after hybridization to a chicken histone H3 + H4 gene probe. B. Tertiary screening from one of the positive signals observed in A. Over 95% of the plaques that were visible gave a strong signal when hybridized with the same chicken histone H3+H4 gene probe.


B
A

Figure 2. Southern blot containing Eco RI-digested λHHG DNA.

All the λCh4A recombinant clones that gave positive hybridization signals on the tertiary screening were grown in liquid, their DNA was isolated and restricted with Eco RI restriction endonuclease and the fragments were separated on a 0.8% agarose gel. Fifteen clones showing positive hybridization with the chicken H3 + H4 histone gene probe are shown. The position of the λ arms after Eco RI digestion is shown at the right side of the figure. Numbers at the top indicate the identity of the λHHG clone in each lane.


55
41
39
34
30
26
24
22
19₃
19₁
17
6
5₄
5₃
2

some cases. The difference in the intensity of hybridization signals may indicate different degrees of homology between the different human genes and the chicken H3 and H4 genes used for hybridization. However, in many cases, the difference seems to correlate simply with the amount of DNA present, since not all the phage preparations gave the same yields. Furthermore, as will be shown later, some of the hybridizing fragments contain more than one histone H3 or H4 gene.

Figure 2 is a composite made from several different autoradiograms showing different times of exposure, as required for different clones to show positive hybridization signals; however, it should be emphasized that the strongest hybridization signals were obtained with clones 5_3, 6, 17, 22, 39, 41 and 55. Intermediate strength signals were obtained with clones 2, 24, 26, 30 and 34, while weak signals were observed in clones 5_4, 19_1 and 19_3. Based on a short exposure, in which only the first seven, strongly positive clones were observed clearly, these seven clones were selected for futher characterization and will subsequently be called λ HHG 5, λHHG 6, λHHG 17, etc. Note that λHHG 5 appears in Figure 2 as 5_3, and is not to be confused with 5_4, a clone that was not further characterized.

Restriction Mapping

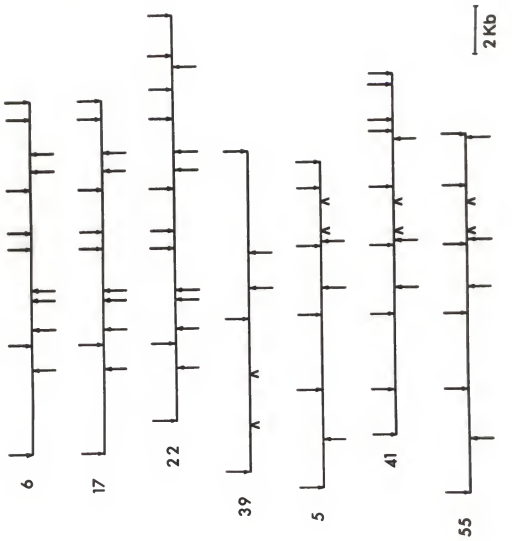
The first step in the characterization of newly isolated clones usually involves both restriction mapping and location, within this restriction map, of the genes of interest. Both goals can be pursued in parallel if appropriate hybridization probes are available, since gels with restriction digests, used to map restriction endonuclease recognition sites, can then be blotted to nitrocellulose and used to hybridize consecutively with up to 4 or 5 different probes, provided the filter is handled carefully, and the probes are properly removed between hybridizations.

Total phage DNAs were mapped with respect to restriction endonucleases Eco RI, Hind III and Bam HI, while selected subclones (see below) were further mapped with respect to several other restriction endonucleases. To construct maps of the λHHG phage, DNA was initially digested with Eco RI, Hind III or Bam HI or with all possible combinations of two of these enzymes. The products of digestion were electrophoretically separated on 0.8% agarose gels. Due to the large size of the inserts and the large number of fragments generated by each of these digestions, it was not possible to construct definitive restriction maps for any of the clones immediately, and the study was then supplemented with partial digestions, as well as the use of different gel systems, such as 0.4%, 1.5% or 3% agarose gels or 5% polyacrylamide gels for the resolution of very small fragments. The most useful tools for constructing accurate restriction maps of the λHHG phage were several histone gene specific probes, which, when hybridized to the previously described digests, allowed the identification of overlapping fragments, derived from different digestions, all of which bore a specific human histone gene.

Figure 3 shows the restriction maps obtained for all seven λHHG clones. The different clones can be grouped into three classes, according to their restriction sites. λHHG 6, λHHG 17 and λHHG 22 form one group with overlapping restriction maps. It should be noted that, for purposes of comparison, the insert in λHHG 22 is displayed in the opposite orientation with respect to the λ arms, so that its similarity with clones λHHG 6 and λHHG 17 is emphasized. Clones λHHG 6 and λHHG 17 appear to be identical and might in fact be two independent isolates of the same recombinant. This could be due to replication of the recombinant during the amplification of the library (207).

Figure 3. Restriction maps of seven λHHG clones.

λHHG clones were mapped with respect to Eco RI (↓), Hind III (↓) and Bam HI (^) restriction endonucleases. The bar at the bottom of the figure indicates the scale. Note that, for reasons of comparison, the insert present in clone λHHG 22 is displayed in an opposite orientation with respect to the λ arms, as compared with all other λHHG clones.


6
17
22
39
5
41
55
2 Kb

Further analysis of subclones derived from these clones, performed by Richard Rickles in our laboratory, has shown that clones λHHG 6 and λHHG 17 are also identical with respect to several other restriction endonucleases; however, digestion of DNA from a subclone obtained from λ HHG 22 (pTT 915) with restriction endonuclease Sac II gives rise to a pattern that differs from the Sac II pattern obtained from the equivalent subclones of λHHG 6 (pSX 915) or λHHG 17 (pST 519) (data not shown).

Clone λHHG 39 stands in a class by itself; no one of the other six λHHG clones shares a restriction pattern with it.

Clones λHHG 5, λHHG 41 and λHHG 55 gave a third class of restriction patterns. Their maps form a set of overlapping DNA fragments, but all three appear to be independent isolates, since the junctions with the λ arms are located at different positions in each of the clones. Again, detailed restriction analysis of the corresponding subclones isolated from these phage showed some minor differences in the Alu I pattern obtained from a subclone of clone λHHG 5 (pFV 911), as compared with those obtained from either clone λHHG 41 (pFO 536) or λHHG 55 (pFF 201). Also, in work done by Aleida Leza and Dr. Farhad Marasni, a subclone derived from λHHG 55 (pFF 428) showed a Pst I site not present in the equivalent subclone derived from λHHG 41 (pFO 108) (not shown).

The results described above suggest that these clones might represent independent members of a family of related clusters of human histone genes, which in turn could form part of a larger "repeat," although this repeat is clearly not as simple as those observed for the histone genes of sea urchins or D. melanogaster, as will be shown later. The observation of differences between 3 independent, similar clones, makes unlikely the simplistic notion that two closely related, but slightly different clones

might represent the two alleles derived from diploid fetal liver cells, one arrangement being derived from each of the two parents. Furthermore, preliminary data obtained in our laboratory on the genomic organization of human histone genes seems to indicate that these clones do indeed represent major repeats of histone genes in humans, although the results are not conclusive yet, as will be discussed in a later section. Genomic blots have also shown several minor arrangements, a fact that emphasizes the necessity to further analyze other isolates from the genomic library.

Histone Coding Regions

A. cDNA:

Several radioactive probes were used to identify and localize the human histone genes present in the λHHG clones. As previously described, the clones were isolated by using a chicken histone probe containing genes coding for histones H3 and H4. Because no isolation of human histone genes had been reported in the literature at the time these experiments were performed, most of the identification and localization of the histone genes within the λHHG clones was done using heterologous probes, namely, from chicken and from two different species of sea urchins. This approach is reasonable, considering that histone proteins are among those most conserved throughout evolution, a fact that suggests that the DNA sequences might be similarly conserved.

On the other hand, many third base differences (wobbling), or differences in non-coding regions could occur, which would still give rise to the same protein, while complicating the hybridization studies. In this respect, it is interesting to note that Dr. Alex Lichtler in our laboratory identified at least seven different subspecies of H4 mRNA, which differ from one another in primary structure (80,204). Furthermore,

experiments done in collaboration with Dr. Lichtler, and using the λHHG clones, have shown that the different H4 mRNAs do not share significantly long stretches of homologous sequences, since digestion of their hybrids with S_1 nuclease did not give rise to intermediate size bands on gels: a different subspecies of H4 mRNA was protected by each different H4 gene, while all other mRNA subspecies were degraded to small fragments (204). In the same context, Kunkel and Weinberg (228) have reported the lack of hybridization between clones containing the early histone genes from the sea urchin Strongylocentrotus purpuratus, and late histone mRNA sequences extracted from the homologous organism.

Although these two examples both involved DNA/RNA hybridization, which is in general more sensitive to mismatches than DNA/DNA hybridization (229), one of the first goals on characterizing the HHG phage with respect to histone coding regions was to show that they would hybridize efficiently to a homologous probe containing histone sequences. Due to the above considerations, hybridization with histone-enriched mRNA directly was thought to be less likely to succeed than a DNA/DNA hybridization. We prepared a cDNA to mRNA that had been enriched in histone mRNA sequences (as shown by in vitro translation). cDNA to 7-11 S RNA from HeLa S_3 cells was prepared as described in Materials and Methods and was then hybridized to a nitrocellulose filter containing electrophoretically fractionated Eco RI-digested λHHG DNA. Figure 4B shows the results obtained from this hybridization. Several of the fragments obtained from the insert present in every one of the λ phage hybridized with the cDNA probe. Analysis of the sizes, as well as the number, of bands hybridizing to the cDNA probe, as compared with the chicken H3 plus H4 probe (Figure 2), indicates some interesting observations:

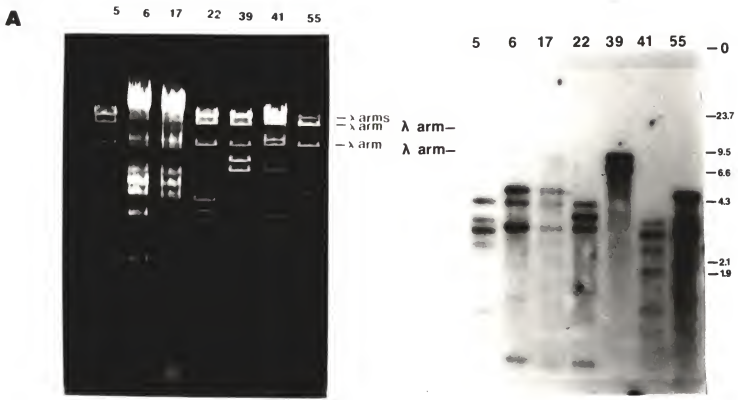
1. The two probes did not always hybridize with a specific fragment with the same relative intensity. (Notice, for example, the three major hybridizing bands from λHHG 22, of sizes between 3.0 and 4.0 Kb.)
2. Some bands show strong hybridization with the cDNA probe, while they do not hybridize at all with the chicken H3 plus H4 probe (for example, the 6.5 Kb band from λHHG 39).
3. Several bands hybridized with the cDNA probe with relatively low intensity, but did not show positive hybridization signals with the chicken histone probe.
4. No band that hybridized with the chicken H3 plus H4 probe failed to hybridize with the cDNA probe.
5. Not all the bands show hybridization to the cDNA probe (compare Figures 4A and 4B); particularly, the λ arms did not hybridize, as well as several of the bands from clones λHHG 5, λHHG 6, λHHG 17 and λHHG 22.

The results were interpreted to mean that fragments containing histone genes gave the strongest signals, and some fragments contained histone genes other than H3 or H4. Minor hybridization signals were attributed to hybridization of the DNA with other RNA coding genes, interspersed with the histone genes in the human genome. Although these conclusions were not the only possible explanation of the observed results, they were later confirmed to be correct (see below).

Specific histone gene coding regions were originally assigned by hybridization of Southern blots containing restriction endonuclease digests of λHHG DNA to several heterologous, nick-translated histone DNA probes.

Figure 4. cDNA hybridization.

A. Ethidium bromide stain of a gel containing electrophoretically separated Eco RI restriction digests of λHHG DNA. The position of the λ arms is indicated at the right side of the figure. B. Hybridization of λHHG clones to histone-enriched cDNA. ^{32}P-labeled cDNA to 7-11S polysomal RNA from S phase HeLa cells was prepared as described in Materials and Methods. Hybridization to a Southern blot containing Eco RI-digested λHHG DNA was for 20 hours. Numbers on the right indicate the size (in Kbp) of λHind III markers electrophoresed in the same gel.


A
5
6
17
22
39
41
55
λ arms
λ arm
λ arm
λ arm–
λ arm–
5
6
17
22
39
41
55
–0
–23.7
–9.5
–6.6
–4.3
–2.1
–1.9

B. H4:

A chicken H4 histone gene probe was prepared by digesting p2.6H DNA with Sac II and Sma I restriction endonucleases (Figure 5). The 1.1 Kb fragment containing the H4 gene was isolated from a low gelling temperature agarose gel, nick-translated and hybridized to several nitrocellulose filters containing different digests of λHHG DNAs. Figure 6 shows the results obtained when this probe was hybridized to Eco RI digests of λHHG DNAs. All seven of the λ phage under study showed hybridization with this H4 histone gene probe. As expected from the restriction maps (although no maps were available at the time the experiment was done), λHHG 5, λHHG 41 and λHHG 55 all contain a similar DNA band, 3.0 Kb long, that hybridized to the H4 probe, while λHHG 6, λHHG 17 and λHHG 22 all share a hybridizing band of 4.0 Kb in length. Interestingly, these last three clones all showed hybridization with yet another band, with a length of 4.7 Kb in λHHG 6 and λHHG 17, and of 3.5 Kb in λHHG 22, indicating that either the H4 gene is split by an Eco RI site, or there are two H4 genes on each of these clones.

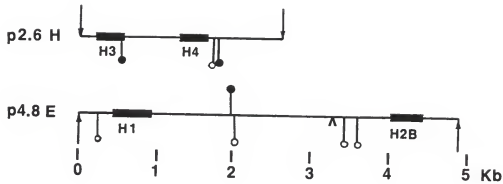
Analysis of the restriction maps shown in Figure 3 indicates that the hybridizing bands are not in contiguous positions, thus making it clear that these three clones contain two independent H4 genes each.

Hybridization of the H4 probe with λHHG 39 DNA produces a positive hybridization signal with the same Eco RI fragment that had previously been shown to hybridize with the H3 plus H4 chicken histone gene probe.

C. H3:

A 570 bp fragment containing a chicken H3 histone gene was prepared by digestion of p2.6H DNA with restriction endonucleases Hind III and Sac II. The results obtained when this probe was hybridized to Eco RI digests

Figure 5. Restriction map of chicken genomic histone gene clones. These clones were generously provided by Drs. Susan Clark and Julian Wells, and were further mapped in our laboratory. The positions of histone coding regions were determined by hybrid selection-in vitro translation, as well as by direct sequencing.


p2.6 H
H3
H4
p4.8 E
H1
H2B
0
1
2
3
4
5 Kb

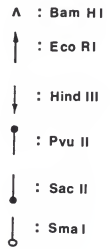
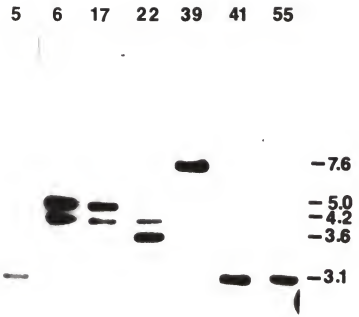

A : Bam HI
: Eco RI
: Hind III
: Pvu II
: Sac II
: Sma I

Figure 6: Hybridization of λHHG clones to probes for specific histone genes. H4 hybridization.

Hybridization of a Southern blot containing Eco RI-digested λ HHG DNA with a chicken H4 probe. Numbers at the right refer to the size (in Kilobase pairs) of hybridizing bands present in each λHHG clone as determined by the migration of λHind III markers electrophoresed in parallel lanes of the same gels. Numbers at the top refer to the DNA present on the nitrocellulose filter.


5
6
17
22
39
41
55
7.6
5.0
4.2
3.6
3.1

of λHHG DNA are shown in Figure 7. All the clones, except λHHG 39, contain histone H3 genes. The H3 probe hybridizes with a band of 4.2 Kb present in clones λHHG 5 and λHHG 55. As shown in Figure 3, this fragment is at the left end of the insert in both of these clones, while the corresponding Eco RI fragment from λHHG 41, the other member of the group, is only 1.9 Kb long; this band still hybridized with the H3 probe. This, plus hybridization data obtained with Eco RI/Hind III double digests (not shown), locates the H3 gene at the right-most end of this Eco RI fragment.

Again, λHHG 6, λHHG 17 and λHHG 22 showed two bands hybridizing with this probe, one of which (4.7 Kb in the case of λHHG 6 and λHHG 17, and 3.5 Kb in the case of λHHG 22) also shows hybridization with the H4 probe (see Figure 6).

D. H2B:

The location of H2B coding regions was determined by hybridization with probes isolated from sea urchin, as well as from chicken. A 1.45 Kb HhaI fragment containing the H2B gene from the plasmid pCO2, which contains an entire histone gene repeat from the sea urchin Strongylocentrotus purpuratus (88), was used first as a probe for the H2B gene. Figure 8A shows that this DNA hybridizes very strongly with clones λ HHG 39 and λHHG 55. Note that no DNA from clone λHHG 5 was available at the time this experiment was performed, and thus, its hybridization with the H2B probe from sea urchin is not shown in Figure 8A. Clones λHHG 17 and λHHG 41 showed weak signals on hybridization. However, these signals were not observed when other H2B probes were used (see below) and were not observed reproducibly when the phage DNA was cut with other restriction endonucleases (not shown).

Figure 7: Hybridization of λHHG clones to probes for specific histone genes. H3 hybridization.

Hybridization of a Southern blot containing Eco RI-digested λHHG DNA with a chicken H3 probe. Numbers at the right refer to the size (in Kilobase pairs) of hybridizing bands present in each λHHG clone as determined by the migration of λHind III markers electrophoresed in parallel lanes of the same gels. Numbers at the top refer to the DNA present on the nitrocellulose filter.

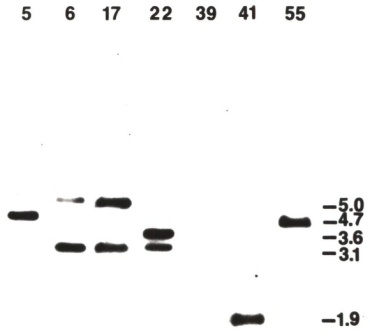
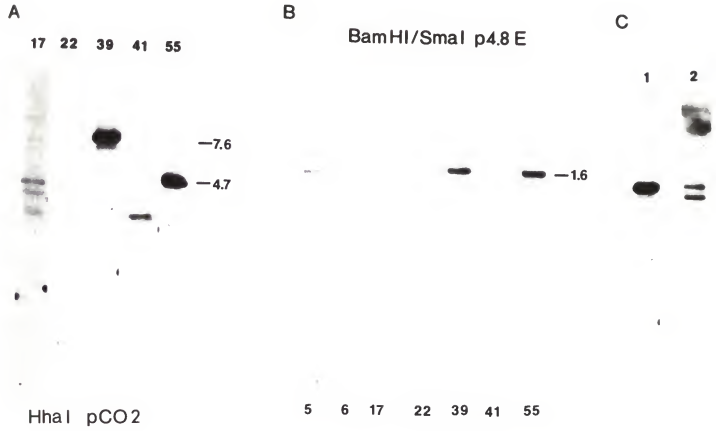

5
6
17
22
39
41
55
5.0
4.7
3.6
3.1
1.9

Figure 8: Hybridization of λHHG clones to probes for specific histone genes. H2B hybridization. A. Hybridization of a Southern blot containing Eco RI-digested λHHG DNA with a sea urchin (S. purpuratus) H2B probe. B. Hybridization of seven identical Southern blots each containing Bam HI/Sma I-digested p4.8E DNA, with radiolabelled λHHG DNA. C. Hybridization of a Southern blot containing Eco RI-digested λHHG 39 DNA with two different radiolabelled H2B probes: Lane 1: The probe was a 1.45 Kb Hha I fragment from pCO2 (same lane shown in A). Lane 2: The probe was nick translated p4.8E.

In the first two panels, designations at the top refer to the DNA present in the nitrocellulose filter, designations at the bottom refer to the nick-translated probe used, and numbers at the right indicate the size (in kilobase pairs) of hybridizing bands present in each λHHG clone as determined by the migration of λ Hind III markers electrophoresed in parallel lanes of the same gels.


A
17 22 39 41 55
—7.6
—4.7
HhaI pCO2
B
BamHI/SmaI p4.8 E
—1.6
5 6 17 22 39 41 55
C
1 2

Because the results with the sea urchin H2B probe were not as clear cut as with the chicken H3 and H4 probes, another chicken probe was used. Clone p4.8E contains both an H2B and an H1 histone gene from chicken (Figure 6) (118). To assay for the presence of these genes in the HHG clones, DNA from p4.8E was digested with Bam HI and Sma I restriction endonucleases, which should separate the two genes: H2B is in a 1.6 Kb fragment, while H1 is in a 1.8 Kb fragment. These two enzymes were chosen based on an incorrect restriction map provided to us by Dr. Julian Wells. An Eco RI/ Sma I digest would have given better separation between the two histone coding regions than the Bam HI/ Sma I digest used; however, the results of this experiment were easily analyzed once an accurate restriction map of p4.8 E was constructed. This DNA was run in seven separate lanes of a 1.5% agarose gel. The DNA was transferred to nitrocellulose filters according to the method of Southern (215), and the filters were then cut in strips that were individually hybridized with nick-translated λHHG DNAs.

The results, shown in Figure 8B, confirmed those obtained with the sea urchin probe: only λHHG 39 and λHHG 55 contain H2B genes. Also, the results show the presence of an H2B gene in λHHG 5, which was not included in the experiment with the sea urchin H2B probe.

Taken together, these results indicate that λHHG 55 contains an H2B gene at the extreme left end of the insert. A similar fragment is also present in λHHG 5, but not in λHHG 41, which does not appear to hybridize with the H2B probes.

Clone λHHG 39 also contains an H2B gene, as evidenced by its hybridization with both probes. However, when the opposite experiment was performed, i.e., nick-translated total insert from p4.8E was hybridized to a

blot containing Eco RI-digested λHHG 39, both of the bands derived from the insert in λHHG 39 hybridized with the probe (Figure 8C). Initially, this result was interpreted to mean that λHHG 39 contained an H1 gene; however, the results in Figure 8B indicated no hybridization between λHHG 39 and the 1.8 Kb band of p4.8E containing the chicken H1 gene.

The results shown in Figure 8C were then reinterpreted to mean that the H2B gene in clone λHHG 39 is split by Eco RI digestion, leaving the more conserved region of the gene in the 7.6 Kb fragment (which hybridizes preferentially with the chicken probe, but not with the H2B probe from sea urchin). Analysis of published sequences for H2B proteins from other species indicates that an Eco RI site might exist in the H2B gene between the sequences encoding for amino acids 90 and 91 (101). Furthermore, it has been found that the amino terminus of the H2B protein is highly variable, while the carboxy terminus is evolutionarily very well conserved (101), a fact that agrees well with the proposed location for the H2B gene in λHHG 39. The presence of the H2B gene in λHHG 39 within 300 bp from the Eco RI site has been confirmed by hybridization experiments performed by Nadine Carozzi and by Keith Prokopp. The presence of H2B sequences in both Eco RI fragments of the λHHG 39 insert has also been shown by hybrid selection-in vitro translation studies performed by Dr. Farhad Marashi. More recently, Keith Prokopp has sequenced part of the H2B gene present in λHHG 39, and his results have confirmed the location of the gene around the Eco RI site, with the more conserved region of the protein encoded in the 7.6 Kb fragment.

E. H2A and H1:

Genes coding for histone H2A proved to be the most difficult to identify. Despite the use of H2A-specific probes derived from two

different sea urchins, Strongylocentrotus purpuratus (the 0.3 Kb Eco RI/Hha I fragment from pCO2) and Echinus esculentus (clone pTS 323), no consistent and/or reproducible hybridization was obtained. The difficulties encountered in using sea urchin probes, both with H2A and H2B are probably related to the large evolutionary span occurring between sea urchins and humans, a problem which is less apparent when using chicken probes. The assignment of H2A coding regions in the λHHG phage was then based solely on hybrid selection-in vitro translation data obtained by Dr. Farhad Marashi.

Hybrid selection-in vitro translation experiments also indicated the absence of histone H1 genes in the λHHG clones. Histone H1 is the least evolutionarily conserved of all the histone proteins; however, hybrid selection-in vitro translation of HeLa histone H1 mRNA had been successfully used in our laboratory to detect the heterologous chicken H1 gene present in p4.8E. The ability to obtain positive hybrid formation between the chicken H1 DNA and human RNA suggests that a negative result in the homologous experiment, i.e., λHHG DNA hybridized with RNA from HeLa S_3, indicates the absence of H1 coding regions in the λHHG phage.

F. DNA Sequencing:

Perhaps the most definitive evidence for the presence of histone genes in the λHHG clones has come from direct DNA sequencing data, obtained by Drs. Terry Van Dyke and Mark Plumb for two H3 genes (the ones on the right side of λHHG 17 and in λHHG 41, respectively), by Keith Prokkopp and Dr. Farhad Marashi for two H2B genes (the ones in λHHG 39 and λHHG 55, respectively) and by myself for an H4 gene (the one in the λHHG 41).

The 3.1 Kb Eco RI fragment containing the H4 gene present in clone λ HHG 41 was subcloned into the Eco RI site of pBR 322 (see below), and this

clone, designated pFO 108, was mapped by Aleida Leza and by myself with respect to a series of different restriction endonuclease sites. For reasons to be explained in a later section, a subclone of pFO 108 was constructed that contained only the 1.8 Kb Eco RI/Hind III fragment from pFO 108. This new clone was named pFO 108A, and hybridization data obtained by Aleida Leza, as well as preliminary in vitro transcription data obtained by myself had suggested that the H4 gene was located predominantly within a 317 bp Sac II fragment, with its 5' end probably located within the adjacent Eco RI/Sac II fragment.

To characterize further the H4 gene present in pFO 108A, the 317 bp Sac II fragment and the 408 bp Eco RI/Sac II fragment were sequenced. For this, pFO 108A DNA was digested with both Eco RI and Sac II restriction endonucleases and the 5' ends of the fragments were labeled by kinasing in the presence of [γ-^{32}P]-ATP. The labeled strands were then denatured and separated on a 5% polyacrylamide gel (Figure 9) and individual strands were excised from the gel and purified by electroelution. It was noted that the faster migrating strand of the 408 bp Eco RI/Sac II fragment was labeled to a much greater extent than the other strands of interest. This phenomenon is very reproducible, and is most likely due to the fact that the end at which this strand is labeled is provided by Eco RI digestion, which leaves a 5'-AATT overhang that is much more readily accessible to BAP and/or kinase than the 3'- overhanging end produced by Sac II digestion.

Figure 10 shows the extent to which each strand was sequenced, as well as the number of times each strand was sequenced. Most of the areas shown were sequenced at least twice.

Figure 9: Strand separation gel.

pFO 108A DNA was restricted with Eco RI and Sac II restriction endonucleases, labelled by kinase in the presence of [γ-^{32}P] ATP. The sample was denatured and electrophoresed on a 5% polyacrylamide strand separation gel. The identity of the different bands was determined by excising and re-running each band under denaturing and non-denaturing conditions.

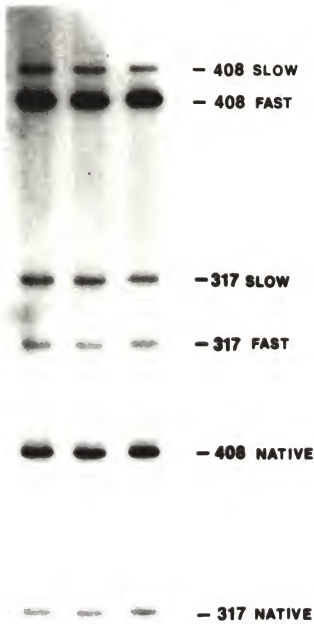

– 408 SLOW
– 408 FAST
–317 SLOW
– 317 FAST
– 408 NATIVE
– 317 NATIVE

Figure 10: Sequencing strategy.

Restriction map of the H4 gene and the flanking regions present in λHHG 41. Horizontal arrows below the map indicate the direction and extent of sequencing information obtained. The number of arrows below each region of the DNA indicates the number of times each fragment was sequenced.

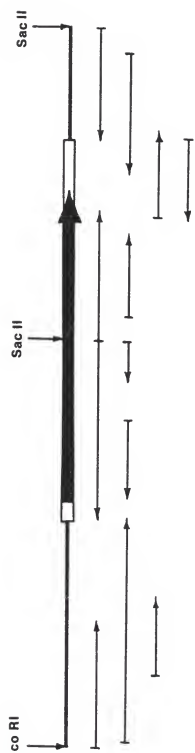

Sac II
Sac II
Eco RI

Figure 11 shows the nucleotide sequence of the H4 gene present in pFO 108A, and its flanking regions. These sequences indicate that pFO 108A DNA has indeed the capacity to code for an H4 protein, whose amino acid sequence does not differ at all from that obtained for other H4 proteins, such as calf thymus (Figures 11 and 12). Furthermore, the results confirm that in humans, as well as in other species studied, histone genes (or at least this particular H4 gene) lack any intervening sequences. This finding is tentatively extended to the non-coding region of the H4 mRNA, based on what is found in other histone gene systems, which indicated that the 3'end of the mRNA is usually located at the ACCA motif immediately following the T-hyphenated dyad symmetry shown in Figure 11 (54). This fact, plus the location of a TATA box slightly upstream from the AUG initiation codon, indicates that the maximum size of the mRNA that could be encoded in pFO 108A is approximately 400 nucleotides, which agrees with previous estimates for the size of a mature H4 mRNA (81).

The protein coding region of pFO 108A has a G+C content of 60.7%, which is similar to the G+C content found in sea urchin histone genes (84-88). Maybe the most interesting observation is that, although the H4 proteins coded by the different genes shown in Figure 12 are identical in their amino acid sequence, the codon usage is not conserved, and a great degree of variability exists in the third base of many codons, even when sequences from two different human histone H4 clones are compared (Figure 12). Nevertheless, the codon usage is not random, as has been previously observed by Turner and Woodland (230).

Sequences obtained for other histone genes present in the λHHG phage will not be presented, since they are only partial sequences, and the experiments were performed by other investigators in our laboratory.

Figure 11: Nucleotide sequence data for a complete H4 gene and its flanking regions.

Sequences were obtained by the Maxam and Gilbert (220) method, using the strategy depicted in Figure 10. Capital letters indicate residues that were consistently determined in a given position. Small letters indicate residues that were not completely clear and/or gave different results in separate experiments. An "N" indicates an undetermined nucleotide.

The boxes at the 5' end of the gene indicate the location of the two tandem "CAAT" boxes, while the Sox just preceding the 3' end of the mRNA coding region indicates the T hyphenated dyad symmetry. The box located just past these nucleotides indicates the histone-related purine box.

Wavy underlines at the 5' flanking region indicate, in order, towards the gene: 1.) twenty-one base pairs of purines, 2.) histone related GTCC motif, and 3.) TATA box.

Horizontal arrows indicate short direct repeats.

-240 -230 -220 -210 -200 -190
AATTC TCCCG GGGAC CGTTG CGTAG GCGTT AAAAA AAAAA AAGAG TGAGA GAGGG ACTGA

-180 -170 -160 -150 -140 -130
GCAGA GTGGA GGAGG AGGGA GAGGA AAACA GAAAA GAAAT GACGA AATGT CGAGA GGGCG

-120 -110 -100 -90 -80 -70
GGGAC AATTG AGAAC GCTTC CCGCC GGCGC GCTTT CGGTT TtCAA TCTGG TCCGA TAtCt

-60 -50 -40 -30 -20 -10
CtGTA TATtA CGGGG AAGaC GGtGa CGCtC CGatC GaNcN Nctat CGGGC TCCtG CGGTC

0
ATG TCC GGC tGt GGa aAG GGC GGA AAG GGC TTA GGC AAA GGT GGC GCT AAG CGC
Met$_0$Ser$_1$Gly Arg Gly Lys$_5$Gly Gly Lys Gly Leu$_{10}$Gly Lys Gly Gly Ala$_{15}$Lys Arg

CAC CGC AAG GTC TTG AGA GAC AAC ATT CaG GGC ATC ACC aAG CCT GCC aTT CGG
His Arg Lys$_{20}$Val Leu Arg Asp Asn$_{25}$Ile Gln Gly Ile Thr$_{30}$Lys Pro Ala Ile Arg$_{35}$

CGT NTA GCT CGG CGT GGC GGC GTT AAG CGG ATC TCT GGC CTC ATT TAC GAG GAG
Arg Leu Ala Arg Arg$_{40}$Gly Gly Val Lys Arg$_{45}$Ile Ser Gly Leu Ile Tyr$_{50}$Glu Glu

ACC CGC GGT GTG CTG AAa GTG TTC TTG GAG AAT GTG ATT CGG GAC GCA GTC ACC
Thr Arg$_{55}$Gly Val Leu Lys Val$_{60}$Phe Leu Glu Asn Val$_{65}$Ile Arg Asp Ala Val$_{70}$Thr

TAC ACC GAG CAC GCC AAG CGC AAG ACC GTC ACA GCC ATG GAT GTG GTG TAC GCG
Tyr Thr Glu His$_{75}$Ala Lys Arg Lys Thr$_{80}$Val Thr Ala Met Asp$_{85}$Val Val Tyr Ala

CTC AAG CGN CAG GGG AGN aCC CtC TAC GGC TTC GGA GGC TAG GCCGC CGCTC
Leu Lys Arg Gln Gly Arg$_{90}$Thr Leu Tyr$_{95}$Gly Phe Gly$_{100}$Gly$_{102}$Stop

mRNA 3' end

CAGCT TTGCA CGTTT CGATC CCAAA GGCCC TTTTT GGGCC GACCA CTTGC TCAtC CTGAG

GAGTT GGACA CTTGA CTGCG TAAAG TGCAA CAGTA ACGAT GTTGG AAGGT AACTT TGGCA

GTGGG GCGAC AATCG GATCT GAAGT TAACG GAAAG acata accgc

Figure 12: H4 sequence comparison for different organisms.

The nucleotide sequence of the H4 coding region of λHHG 41 (bottom line, capital letters) is compared with that of the sea urchin Strongylocentrotus purpuratus (pSp2) (224), sea urchin Psammechinus miliaris (h19 and h22) (93,98), newt Notophthalamus viridescens (Nv 51) (245), frog Xenopus borealis (pc XbH4W1) (230), frog Xenopus laevis (pc Xl H4W2, Xl-hi-1 and p Xlch4) (230,246), mouse (mus-hi-1) (120) and human (pHu4A) (123).

```
pSp 2     tca ggt cga gga aaa gga gga aag gga ctc gga aag ggt ggt gcc
h 19      tca ggt cga gga aaa gga gga aag gga ctc gga aag ggc ggt gcc
h 22      tca ggc cgt ggt aaa gga ggc aag ggg ctc gga aag gga ggc gcc
Nv 51                                                     ggg gct
pcXbH4W1  tct gga aga ggc aag gga gga aag ggt ctg ggg aaa gga ggc gct
pcXlH4W2  tct gga aga ggc aag ggc gga aag ggt ctg ggc aaa gga ggc gct
Xl-hi-1   tct gga cgc ggc aaa gga gga aaa gga ctg ggg aaa gga ggc gcc
pXlch 4                       gga aag ggt ctg ggc aaa gga ggc gcc
mus-hi-1  tct ggc aga gga aag ggt gga aag ggt cta ggc aag ggt ggc gcc
pHu 4A    tct ggc cgc ggc aaa ggc ggg aag ggc ctt ggc aaa ggc ggc gct
pFO 108   TCC GGC tGt GGa aAG GGC GGA AAG GGC TTA GGC AAA GGT GGC GCT
          Ser Gly Arg Gly Lys Gly Gly Lys Gly Leu Gly Lys Gly Gly Ala
             1               5                  10                  15
```

```
aaa cgt cat cgc aag gtt cta cga gat aac atc caa ggc atc acc aag cct gca
aaa cgt cat cgc aag gtt cta cga gac aac atc caa ggc atc acc aag cct gca
aag cgt cat cgc aag gtc cta cga gac aac atc cag ggc atc acc aag cct gca
aag cgg cac agg aag gtg ctc ccN gac aac atc cag ggc atc acc aag cct gct
aag cgc cac agg aag gtg ctg cgg gat aac atc caa ggc atc act aag ccc gcc
aag cgc cac agg aag gtg ctg cgg gat aac atc cag ggc atc acc aag ccc gcc
aag cgg cac agg aag gtg ctt agg gac aac atc cag ggc atc acc aag cct gcc
aag cgc cac agg aag gtg ctg cgg gat aac atc cag ggc atc acc aag ccc gcc
aag cgc cat cgc aaa gtc ttg cgt gac aac atc cag ggt atc acc aag ccc gcc
aag cgc cac cgt aaa gta ctg cgc gac aat atc cat ggc atc acc
AAG CGC CAC CGC AAG GTC TTG AGA GAC AAC ATT CAG GGC ATC ACC AAG CCT GCC
Lys Arg His Arg Lys Val Leu Arg Asp Asn Ile Gln Gly Ile Thr Lys Pro Ala
                  20                  25                  30
```

```
atc cgt cga ctN gct aga agg gga ggt gtc aag agg atc tct ggt ctc atc tac
atc cgt cga ctt gct aga agg gga ggt gtc aag agg atc tct ggt ctc atc tac
atc cgc cga ctc                                  ga atc tct ggt ctt atc tac
atc gNN cgN ctg gcg cgc cNt gga gga gtc aag cgc atc tcc ggc ctc atc tac
atc cgc cgt ctg gcc cgc aga ggt gga gtt aag cgc atc tct ggc ctc atc tac
atc cgc cgc ctg gca cgc aga ggg gga gtc aag cgc atc tcc ggc ctc atc tac
atc cgc cgc cta gca cgg aga ggg gga gtc aag cgc atc tct ggc ctc att tat
atc cgc cgc cta gcc cgc aga ggg ggt gtc aag cgc atc
atc cgc cgc ctg gct cgg cgc ggt ggg gtc aag cgc atc tcc ggc ctc atc tac

aTT CGG CGT NTA GCT CGG CGT GGC GGC GTT AAG CGG ATC TCT GGC CTC ATT TAC
Ile Arg Arg Leu Ala Arg Arg Gly Gly Val Lys Arg Ile Ser Gly Leu Ile Tyr
      35                  40                  45                  50
```

```
gaa gag aca cgc ggt gta ctg aag gtc ttc ctg gag aat gtc atc cgt gat gca
gaa gag aca cgc ggt gta ctg aag gtc ttc ttg gag aat gtc atc cgt gat gca
gag gag aca cga ggg gtg ctg aag g
gag gag acc cgN gNt gtg ctc aag gtt NNc ctg gag aat gtg atc agN Nac gcg
gag gaa act cgc ggg gtg ctg aaa gtt ttc ctg gag aat gtt atc cgg gac gcc
gag gag act cgc ggg gtg ctg aaa gtg ttc ctg gag aac gtt atc cgg gac gcg
gag gaa act cgt ggg gtc ctc aag gtt ttc cta gag aat gtc atc cgg gac gct

gag gag acc cgt ggt gtg ctg aag gtg ttc ctg gag aac gtc atc cgc gac gca
                                                                gac gcc
GAG GAG ACC CGC GGT GTG CTG AAa GTG TTC TTG GAG AAT GTG ATT CGG GAC GCA
Glu Glu Thr Arg Gly Val Leu Lys Val Phe Leu Glu Asn Val Ile Arg Asp Ala
             55                  60                  65
```

```
gtc acc tac tgc gag cac gct aag cga aag act gtc aca gcc atg gac gtg gtg
gtc acc tac tgc gag cac gcc aag cga aag act gtc aca gcc atg gac gtg gtg


gtc acc tac acc gag cac gcc aag agg aag acc gtg acc gct atg gat gtg gtc
gtc acc tac acc gag cac gcc aag agg aag acc gtc acc gct atg gat gtg gtg
gtc acc tac acc gag cac gcc aag agg aag acc gtt acc gcc atg gat gtg gtg
                                        acc gtt acc gcc atg gat gtg gtg
gtc acc tac acc gag cac ggc aag cgc aag acc gtc acc gct atg gat gtg gtg
gtc agc tat aca gag cac gcc aag cgc aag acg gtc acc gcc atg gat gtg gtc
GTC ACC TAC ACC GAG CAC GCC AAG CGC AAG ACC GTC ACA GCC ATG GAT GTG GTG
Val Thr Tyr Thr Glu His Ala Lys Arg Lys Thr Val Thr Ala Met Asp Val Val
  70                  75                  80                  85
```

```
tat gca cta aag agg cag ggt cgt aca ttg tac ggc ttc ggc ggc    pSp 2
tat gca ctg aag agg cag ggt cgt aca ttg tac ggc ttc ggc ggc    h 19
                        ggc cga aca ctg tac ggc ttc ggc ggc    h 22
                                                               Nv 51
tat gct ctc aaa cgt cag ggc cgc act ctc tac ggt ttc gga ggt    pcXbH4W1
tat gct ctg aag cgc caa gga cgc act ctg tac gga ttc gga ggt    pcX1H4W2
tac gct ctc aag cgc cag ggc cgc act ctc tac ggc ttc ggc gga    X1-hi-1
tat gct ctg aag cgc cag gga cgc act ctg tac gga ttc gga ggt    pX1ch 4
tac gct ctc aag cgc cag ggc cgc acc ctc tac ggc ttc gga ggc    mus-hi-1
tac gcg ctc aag cgc cag ggc cgc acc ctc tac ggt ttc ggt ggt    pHu 4A
TAC GCG CTC AAG CGN CAG GGG AGN aCC CtC TAC GGC TTC GGA GGC    pFO 108
Tyr Ala Leu Lys Arg Gln Gly Arg Thr Leu Tyr Gly Phe Gly Gly
          90                  95                 100    102
```

Figure 12 -- continued

From the H4 histone gene sequencing data presented in Figure 11, we have conclusively demonstrated the presence of true histone genes in the λHHG clones; however, it is theoretically possible that this specific gene might be inactive in vivo, due to abnormalities in the regulatory sequences. In conjunction with Dr. Alex Lichtler, we have been able to demonstrate that the H4 gene present in HHG 41 (from which pFO 108A was derived) does completely and strongly protect a single species of full size H4 mRNA obtained from HeLa cells, an observation that suggests that this mRNA might be the in vivo product of the H4 gene present in pFO 108A (204). The sequences present in the flanking regions of the H4 gene will be discussed at a later point.

G. Other Approaches:

Other approaches were taken by several investigators in our laboratory to confirm the presence and location of histone coding regions in λHHG phage. These include the previously mentioned hybrid selection-in vitro translation experiments performed by Dr. Farhad Marashi, Northern blot analysis performed by Richard Rickles and hybridizations to in vivo-labeled RNAs performed by Drs. Alex Lichtler and Mark Plumb.

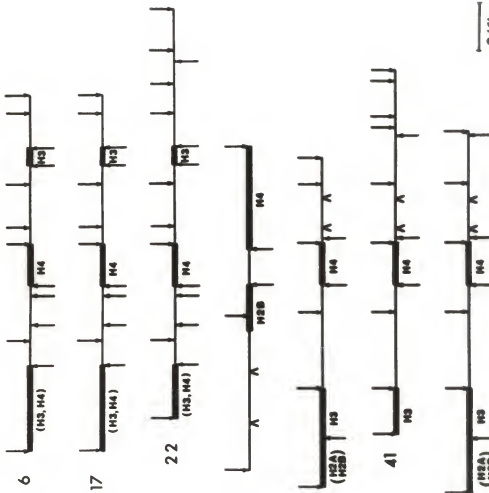
General Features of Human Histone Gene Organization

A. Clustering:

At the time these experiments were initiated, the structure of histone genes in sea urchins and D. melanogaster were known. In both cases, histone genes were found to be clustered and tandemly repeated. The repeats were of about 6 Kb in length in either case (81).

Figure 13 shows a restriction map of each of the seven clones under study. The location of restriction fragments hybridizing to different histone probes is indicated.

Figure 13: Restriction maps of seven λHHG clones, indicating the location of histone coding regions. DNA fragments that hybridize to the different histone gene probes are included, as well as the location of H2A genes, which was determined by hybrid selection-in vitro translation. (↓):Eco RI, (↑):Hind III, (∧):Bam HI recognition sites.


6
(H3, H4)
H4
H3
17
(H3, H4)
H4
H3
22
(H3, H4)
H4
H3
39
H2B
H4
5
(H2A)
(H2B)
H3
H4
41
H3
H4
55
(H2A)
(H2B)
H3
H4
2 Kb

The results clearly indicate that, in humans, histone genes are clustered, but no obvious repeat is apparent. All the clones contained at least two different histone genes adjacent to each other (within a few Kb from one another). However, none of the clones seems to contain a complete set of histone genes, since H1 genes have not been detected in any of these clones. Nevertheless, clones λHHG 5 and λHHG 55 each contain a whole set of core histone genes, that is, H2A, H2B, H3 and H4. Within the limits of our analysis, only one copy of each one of these genes is present on each one of these clones.

On the other hand, clones λHHG 6, λHHG 17, and λHHG 22 each contain two genes coding for each of the inner core histones, H3 and H4.

These patterns of arrangement preclude the existence in humans of simple repeats like those found in sea urchins and Drosophila, where each repeat contains one of each of the five histone genes. At the same time our work on the organization of human histone genes was being done, reports appeared in the literature, describing the same type of organization of the histone genes, based on clusters but with no simple repeat in other organisms, such as yeast (105), mouse (120,121), Xenopus (111,112), chicken (114-118) and the newt Notophthalamus viridescens (106,109). Of course, the possibility that tandem repeats do exist in any one of these organisms or in humans cannot be formally excluded based on the available information. It is also conceivable that some of the clones described represent incomplete fragments from a larger tandem repeat. These possibilities can be better defined with a detailed study on the genomic organization of histone genes in human DNA, using the described clones as sources for appropriate probes. This work is currently in progress in our laboratory.

B. Interspersion with other Sequences:

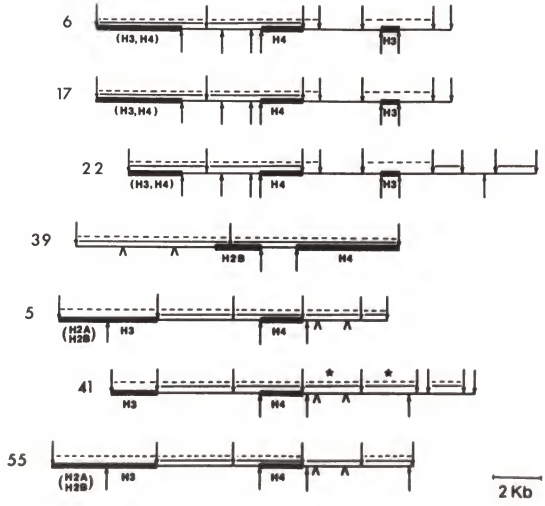
It has been known for several years that the genomes of higher eukaryotes contain, in addition to single copy sequences, large amounts of DNA of moderate to high repetition (5,6). In most eukaryotes, highly repeated sequences approximately 300 nucleotides long have been found interspersed with single copy sequences throughout the genome (231). In humans, the Alu DNA family is predominant among these reiterated sequences, with a repetition frequency of approximately 300,000 per haploid genome (232). Tashima et al. (233) have reported that over 95% of the recombinants present in a human genomic library cloned into λCh 4A hybridize to a probe containing Alu DNA sequences. Sequences related to the Alu DNA family have been found to be transcribed in vitro by RNA polymerase III (234), and they might also be transcribed in vivo, since Alu sequences have been found associated with Hn RNA of human K 562 cells (235) and with poly A^+ polysomal RNA from CCRF-CEM, another human cell line (236).

All the above considerations, plus the observation that a cDNA probe prepared to 7-11 S polysomal RNA from HeLa S_3 cells hybridized with Eco RI fragments derived from λHHG clones to which no histone coding region was assigned (Figure 4A), prompted us to explore further the possibility that Alu DNA sequences were interspersed with the human histone genes present in the λHHG phage.

In these experiments, performed by Aleida Leza, Eco RI digests of λHHG DNA were electrophoretically fractionated, transferred to nitrocellulose and hybridized to nick-translated pCDF 2 (a gift from Dr. S. Weissman). This recombinant plasmid contains a 482 bp insert, most of which (300 bp) corresponds to a member of the Alu family found in the human β-globin gene cluster. The results (not shown) indicated that

Figure 14: Interspersion of histone genes with other transcribed sequences.

The map is identical to that on Figure 13. Full lines above each clone indicate restriction fragments that hybridize to an Alu DNA probe; broken lines above each clone indicate restriction fragments that hybridize with the cDNA to HeLa 7-11 S, S phase polysomal RNA (see Figure 4B). The two fragments denoted with a star (*) migrate as a doublet, and therefore the hybridization results with respect to these two fragments are inconclusive.


6
(H3, H4)
H4
H3
17
(H3, H4)
H4
H3
22
(H3, H4)
H4
H3
39
H2B
H4
5
(H2A)
(H2B)
H3
H4
41
H3
H4
55
(H2A)
(H2B)
H3
H4
2 Kb

several Eco RI fragments hybridized both with the Alu DNA probe and with the cDNA probe. However, several of the fragments hybridized to only one of the two probes (Figure 14). We concluded from these experiments that the histone genes, a family of moderately repeated and clustered genes themselves, are interspersed with members of the Alu DNA family, as well as with other transcribed sequences (202), since some Eco RI fragments hybridize with the cDNA probe, yet do not contain Alu DNA sequences, as shown by their lack of hybridization to pCDF 2, and do not contain histone coding sequences either.

Subcloning into pBR 322

Genomic libraries are best constructed in λ vectors (205,207) or in cosmids (237), since these vectors were developed with the specific aim of accommodating a large insert. This allows the whole genome of an organism to be contained within a manageable number of recombinants. As previously mentioned, the human library in λCh 4A used to select histone genes was contained in 8 X 10^5 recombinant phage. However, after isolation, it is usually convenient to transfer the DNA sequences of interest into smaller vectors, such as pBR 322 or other plasmids, so that the DNA can be handled and analyzed in more detail, without the presence of several Kb. of vicinal, but unrelated sequences.

We subcloned the Eco RI fragments derived from λHHG clones into the Eco RI site of pBR 322 with various purposes in mind: 1. smaller vectors containing a limited amount of histone-related DNA would permit more accurate restriction mapping of genes of interest, 2. several histone genes would now be separated into different plasmids, thus facilitating their use as molecular probes of histone gene expression, 3. studies on *in vitro* transcription would be easier to design and interpret if only one

gene is analyzed at a time, 4. pBR 322-derived clones would serve as a first step towards the construction of coding sequences-only probes, necessary for studies on the genomic organization of histone genes, 5. finally, pBR 322-derived clones are easier to grow and handle than λCh 4A ones, thus allowing good yields of clean DNA to be obtained within few days.

The decision to clone the Eco RI fragments was made based on the fact that the library was originally constructed by ligating Eco RI-digested λ Ch 4A arms to human DNA fragments containing Eco RI linkers (207). Consequently, digestion of λHHG clones with Eco RI gives rise to only two λ Ch 4A fragments. Each has one end that can be ligated to Eco RI-digested pBR 322, however the other end contains the λ cohesive end and cannot be ligated to form a circular molecule. If the two arms are allowed to anneal with each other through their cohesive ends, they would form a chimeric molecule 30.5 Kb long, having Eco RI cohesive termini at both ends. Although these molecules could in theory ligate to pBR 322 to form a circular molecule, such a plasmid could not be replicated in E. coli, since the insert would be too long (238). In conclusion, digestion of λ HHG clones with Eco RI, followed immediately by ligation to calf intestine alkaline phosphatase-treated pBR 322, should give rise to a collection of recombinants containing all the Eco RI fragments derived from λHHG inserts, with virtually no false positive clones.

The experiment, performed as described, gave a large number of recombinants, most of them containing at least one Eco RI fragment derived from the λHHG inserts. Several clones contained more than one Eco RI fragment, and thus it was necessary to screen relatively large numbers of recombinants to obtain the whole collection. This work was done with the collaboration of several post-doctoral fellows in our laboratory.

As a result, essentially all the Eco RI fragments were subcloned into pBR 322, with three exceptions: one fragment from λHHG 17, whose equivalent from λHHG 6 (pSX 919), had been isolated, one small fragment from λHHG 22, and one fragment from λHHG 41 (see Figure 15). Due to the small size of these two fragments, and, in the case of the fragment from λHHG 41, its relatively large distance from any known histone gene, it was considered unnecessary to further pursue subcloning them.

Figure 15 shows again the restriction maps of all seven λHHG phage. A number aoove each Eco RI fragment identifies the subclone containing this specific piece of DNA. Nomenclature was chosen to indicate the origin of each subclone with respect to the λHHG phage. Names were made out of two capital letters, which refer to the number of the λHHG clone from which the subclone is derived. This way, subclones from λHHG 5 (FIVE) are called pFV, from λHHG 6 (SIX) are called pSX, from λHHG 17 (SEVENTEEN), pST, from λHHG 22, pTT, from λHHG 39, pTN, from λHHG 41, pFO, and finally, from λHHG 55, pFF.

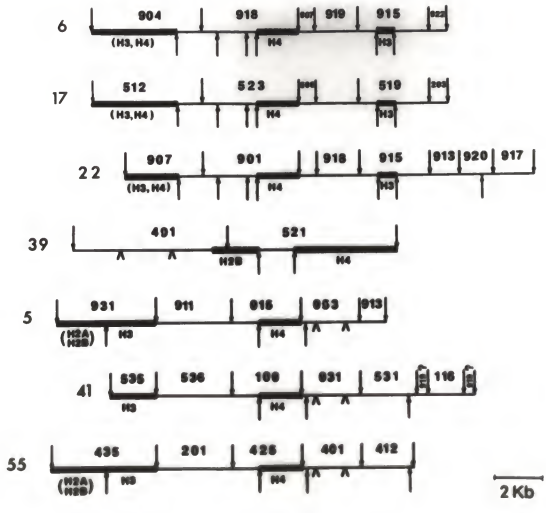
In vitro Transcription

A. Considerations in the Selection of an in vitro Transcription System:

Two independent in vitro transcription systems have been described in the literature and used to effect the accurate transcription by RNA polymerase II of several eukaryotic and viral genes. The systems are a crude S-100 cytoplasmic preparation described by Weil et al. (143) and a total cellular ammonium sulfate cut described by Manley et al. (144).

The S-100 cytoplasmic extract contains one or several factors required for specific initiation of transcription on isolated DNA templates. Since the extract is obtained from the cytoplasmic fraction after hypotonic lysis of the cells, it is assumed that these factors have

Figure 15: Designations of pBR 322 derived subclones.
The map is identical to that on Figure 13. Numbers above each Eco RI fragment identify the subclone containing this specific fragment of DNA. See text for complete nomenclature of the subclones.


6
904
918
919
915
(H3, H4)
H4
H3
17
512
523
519
(H3, H4)
H4
H3
22
907
901
918
915
913
920
917
(H3, H4)
H4
H3
39
491
521
H2B
H4
5
931
911
916
913
(H2A, H2B)
H3
H4
41
536
536
100
931
531
116
H3
H4
55
435
201
426
401
412
(H2A, H2B)
H3
H4
2 Kb

in fact leached from the nucleus. However, no RNA polymerase II activity has been detected to leave the nucleus under the conditions used in this protocol (143), and thus, the system is completely dependent on the addition of exogenous RNA polymerase II, as well as an appropriate DNA template bearing an RNA polymerase II promoter (143). This extract accurately initiates transcription of most RNA polymerase II-dependent genes when supplemented with exogenous RNA polymerase II preparations obtained from different sources, although differences in the rate of transcription and amount of RNA accumulated have been reported (143), and tentatively attributed to differences in the strength of the promoters (143).

The cellular extract described by Manley et al. (144) is a broad ammonium sulfate cut, containing both nuclear and cytoplasmic components. This system does not require the addition of exogenous RNA polymerase II. The cellular DNA is completely removed from the lysate; thus transcription is entirely dependent upon the addition of exogenous DNA templates (144).

Both systems, in appropriate ionic environments, and at the correct ratio of template to extract will accurately initiate transcription of most, but not all genes transcribed in vivo by RNA polymerase II. However, neither of the two systems will accurately terminate transcription of any gene reported to date. This could be due to the lack of appropriate termination factors in both systems or, alternatively, it could be related to the fact that the template DNA, usually cloned sequences, is not in a native conformation with respect to methylation or its association in a chromatin structure. In this respect, it is interesting to note that transcription in vivo has been reported to occur in association with the nuclear matrix (239), a structure not present in either of the in vitro transcription systems described above.

Although not directly addressed in this work, a long range objective of the research in our laboratory is the characterization of factors regulating histone gene expression in vivo. It is my view that this long range project can be best approached by first defining the factors required to obtain transcription of the histone genes in vitro. This in turn involves two independent aspects: 1. determination of those nucleotide sequences in or around a given gene which are required for its transcription, 2. definition of other factors present in the in vitro extracts, that will affect transcription of a competent template.

In this work, we have started the analysis of the DNA sequences required for in vitro transcription of a human H4 gene, and although some progress was made, more work on this project is clearly necessary (see Discussion). The in vitro transcription system chosen for these studies should be amenable to the second part of the project, namely, the characterization of factors other than the template and the RNA polymerase II that might be involved in the transcription of histone genes. The cytoplasmic S-100 extract has the advantage of requiring exogenous RNA polymerase II, and so, if transcriptional activity is lost or restored upon fractionation and reconstitution of the extract, these results can immediately be ascribed to the removal or addition of factors necessary for in vitro transcription. However, being a cytoplasmic extract, this system may lack some of the factors required for the specific recognition of a particular gene or set of genes, because of lack of diffusibility from the nucleus. This potential problem is less likely to be encountered with the whole cell lysate system. On the other hand, the requirement for exogenous RNA polymerase II by the S-100 cytoplasmic system, which, as previously discussed, would facilitate the interpretation of data obtained

by fractionation and reconstitution of the lysate, would introduce an additional variable in the standardization of the system to transcribe the human histone genes.

For these two reasons, plus some preliminary data obtained with both systems (not shown), it was decided that the whole cell lysate was the most appropriate system to use for these studies.

B. Considerations in the Selection of a Template:

When studies with the in vitro transcription system were started, little was known about the exact location or orientation of any histone gene present in the subcones described above. For that reason, and while information was gathered, preliminary in vitro transcription studies were done with pFF 435, a clone containing three histone genes, one each of H2A, H2B and H3 (Figure 13). This arrangement increases the probability that at least one of the genes will have all the required 5' and 3' flanking regions. In retrospect, it is not surprising that results obtained with this template were very confusing and difficult to interpret. At the time, however, mapping and sequencing data were obtained for clones pFO 108 and pST 519, containing an H4 and an H3 gene, respectively. These clones were then chosen for further studies using the in vitro transcription system.

Since the in vitro transcription system does not recognize termination signals (144), it is desirable to be able to cut the DNA template with an enzyme that truncates the gene at a position relatively close to the origin of transcription, so that transcription products terminating at the end of the fragment can be analyzed in tight polyacrylamide-urea gels. This widely used approach is known as a "run-off" transcription assay (141,143-145,152-155). The partial

nucleotide sequence of pST 519 (obtained by Dr. Terry Van Dyke) revealed the presence of a Pvu II site a few nucleotides within the coding region of the H3 gene. The mapping of pFO 108 revealed that part, but not all of the gene was contained within a Sac II fragment, suggesting that Sac II would necessarily cut the gene somewhere within the mRNA coding region (Figure 16A). Both DNAs were used as templates after digestion with the aforementioned enzymes, but no specific run-off transcripts were observed (not shown). However, when pFO 108 DNA was digested with Eco RI, which separates the vector sequences from the insert, α-amanitin sensitive transcription of an RNA species of approximately 2.8 Kb was observed (Figure 17). This is the size expected for a run-off transcript produced from Eco RI-digested pFO 108 DNA.

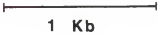
In experiments performed together with Dr. Alex Lichtler, we have shown that the H4 gene present in pFO 108 codes for one of the major subspecies of histone H4 mRNA present in HeLa S_3 cells (204). This fact, plus the fact that α-amanitin sensitive transcription of this gene in the in vitro system was obtained, prompted the selection of this specific subclone for further examination.

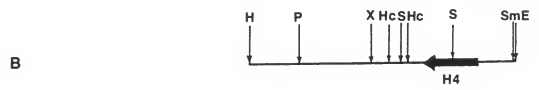
C. Standardization of the in vitro Transcription System:

The original report by Manley et al. (144) clearly shows that the specific transcription of an Adenovirus 2 late gene is dependent on two variables: the concentrations of DNA template and lysate present, as well as the ratio between these two. Furthermore, other reports have shown that different DNA templates require different concentrations of lysate and DNA, as well as different salt conditions to produce accurately initiated transcripts (144,152,155). For this reason, it was necessary to standardize the conditions required for accurate transcription of the human H4 gene present in pFO 108.

Figure 16: Restriction maps of pFO 108 and pFO 108A.

A. Restriction map of clone pFO 108. The black arrow indicates the location and direction of transcription of the H4 histone gene. The white arrow indicates the location and direction of transcription of the putative Alu DNA sequence. B. Restriction map of clone pFO 108A. This clone was constructed by removal of the left side Eco RI/Hind III fragment from pFO 108.





E : Eco RI

Sm : Sma I

H : Hind III

P : Pvu II

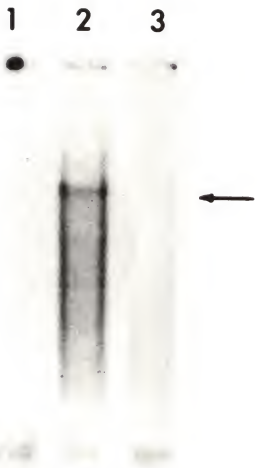
X : Xba I

Hc : Hinc II

S : Sac II

Figure 17: In vitro transcription of Eco RI-digested pFO 108.

In vitro transcripts were analyzed on a 1.5% agarose, 3% formaldehyde gel. Transcripts were then visualized by autoradiography. Lane 1: In vitro transcripts synthesized using Eco RI-digested pBR 322 as a template. Lane 2: In vitro transcripts synthesized using Eco RI-digested pFO 108 as a template. Lane 3: In vitro transcripts synthesized using Eco RI-digested pFO 108 as a template, but in the presence of 2 ug/ml of α-amanitin. The arrow indicates the expected size of a run-off transcript.


1
2
3

The amount of lysate present in a 25 ul reaction greatly influences the appearance of a proper transcript generated from Eco RI-digested pFO 108 DNA at a concentration of 10 ug/ml, as shown in Figure 18. Little 2.8 Kb transcript was obtained when only 10 ul of lysate were present, while a maximum of specific transcription was obtained when 15 ul of lysate (or 60% of the total volume) were used (Lane 2, Figure 18). More lysate (20 ul, or 80% of the total volume) completely abolishes the production of the 2.8 Kb transcript.

DNA concentrations tested included 20, 50 and 100 ug/ml, and were all tested using 15 ul of lysate per 25 ul reaction volume. DNA concentration experiments shown in Figure 19 indicated that optimal transcription was obtained with 50 ug/ml of template DNA. It is clear from the results in Figure 19 that even higher levels of transcription were obtained at a DNA concentration of 100 ug/ml, however, a higher level of background transcription was observed, a fact that would make the interpretation of results more difficult.

Several salt conditions, as well as nucleoside triphosphate concentrations reported in the literature were tested for optimal transcription. It was found that the conditions suggested by BRL (Bethesda Research Laboratories) gave the highest and most reproducible transcription levels, with the minimum of non-specific transcription. Figure 20 shows the effect of the concentration of nucleoside triphosphates on the <u>in vitro</u> transcription of Eco RI-digested pFO 108 DNA. The UTP concentration was maintained at 10 uM, to insure incorporation of radiolabeled [α-^{32}P] UTP. Lanes 1 and 2 show that a much higher level of specific transcription is obtained when 1 mM NTP is used, as opposed to 500 uM. Furthermore, this experiment shows that a

Figure 18: Effect of lysate concentration on the in vitro transcription of Eco RI-digested pFO 108 DNA.

Autoradiogram of a 1.5% agarose, 3% formaldehyde gel showing the transcripts obtained in vitro in a 25 ul reaction, using 20 ug/ml of Eco RI-digested pFO 108 as a template, and varying amounts of whole HeLa cell extract. Lane 1: 10 ul of lysate. Lane 2: 15 ul of lysate. Lane 3: 20 ul of lysate. The arrow indicates the expected size of a run-off transcript.

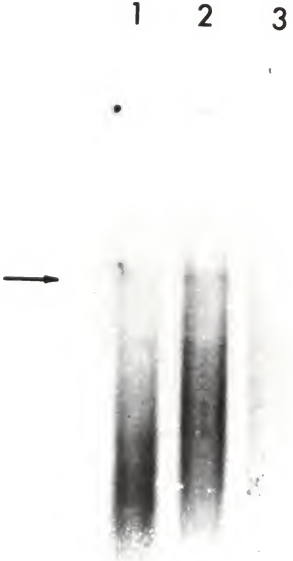

1
2
3

Figure 19: Effect of DNA concentrations on the in vitro transcription of Eco RI-digested pFO 108 DNA.

Autoradiogram of 1.5% agarose, 3% formaldehyde gel showing the transcripts obtained in vitro in a 25 ul reaction, using 15 ul of lysate and varying amounts of Eco RI-digested pFO 108 DNA as a template. Lane 1: no DNA template. Lane 2: 20 ug/ml. Lane 3: 50 ug/ml. Lane 4: 50 ug/ml in the presence of 2 ug/ml of α-amanitin. Lane 5: 100 ug/ml. The arrow indicates the expected size of a run-off transcript.

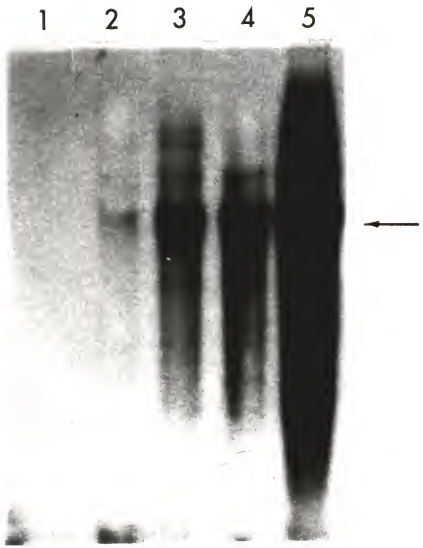
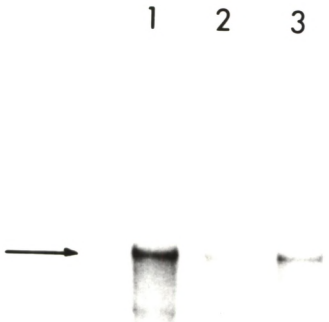

1
2
3
4
5

Figure 20: Effect of nucleoside triphosphate concentrations on the *in vitro* transcription of Eco RI-digested pFO 108 DNA.

Autoradiogram of a 1.5% agarose, 3% formaldehyde gel showing the transcripts obtained *in vitro* in a 25 ul reaction, using Eco RI-digested pFO 108 DNA as template. Lane 1: 1 mM each of ATP, CTP and GTP, with 10 uM UTP. Lane 2: 500 uM each of ATP, CTP and GTP, with 10 uM UTP. Lane 3: Same as in lane 2, but after 45 minutes of reaction, the system was chased for 15 minutes in the presence of 500 uM UTP.

Total reaction time was 65 minutes for all three lanes. The arrow indicates the expected size of a run-off transcript.


1
2
3

fifteen minute chase in the presence of 1 mM UTP increases the production of full size 2.8 Kb transcripts (compare lanes 1 and 3).

Specific conditions used throughout the rest of this work are described in Materials and Methods.

The in vitro transcripts were isolated by a modification of a method developed by Dr. Mark Plumb in our laboratory for isolation of total cellular RNA. This protocol, which involves the addition of three independent ribonuclease inhibitors (SDS, proteinase K and polyvinyl sulfate, see Materials and Methods) was short, reliable and it did not give rise to excessive degradation of the large RNA molecules obtained from the in vitro synthesizing system. In this respect, it should be mentioned that the "run-off assay" previously described, which produces RNA molecules short enough to be analyzed on tight polyacrylamide gels, was not successful. As will be shown later, templates digested with enzymes that would give rise to short run-off transcripts were not suitable substrates under the conditions used. For this reason, extreme care was required in order to maintain the integrity of the long transcripts obtained in vitro.

Finally, and also due to the large size of the transcripts, the RNA transcripts were analyzed in formaldehyde-containing 1.5% agarose gels (222), instead of the traditional 8-10% polyacrylamide-8.3 M urea gels used by other investigators.

D. Detection of Specific Initiation in vitro:

Specific initiation of transcription at the correct 5' terminus of the H4 gene present in pFO 108 was initially assessed by the production of transcripts of the expected size which were sensitive to low concentrations of α-amanitin (2 ug/ml), suggesting transcription by RNA

polymerase II, as expected for histone genes (240). These experiments are described in detail in the next section, however, they only give a rough estimate of the site of initiation of transcription by the in vitro system.

Several methods have been used to map more accurately the site of initiation of transcription by in vitro systems. If the nucleotide sequence of the expected RNA is known, it is possible to predict the size and composition of the oligonucleotides produced by any given ribonuclease digestion. In this case, the ^{32}P-labelled expected to represent the correct transcript can be excised from a gel, digested with a ribonuclease, such as T_2, T_1 or A, and the oligonucleotides can be analyzed by two dimensional (2-D) fingerprint analysis (171,175). This type of analysis of the in vitro transcripts of pFO 108 DNA was not possible because of the large size of the transcripts, which would most likely have produced a 2-D pattern too complex to be compared with the available 2-D patterns obtained by Dr. Alex Lichtler for the RNA encoded by pFO 108 (80,204). Furthermore, the extra sequences at the 3'end of the in vitro products (see below) would certainly produce extra oligonucleotides not present in the in vivo RNA. Alternatively, only the 5' cap structure of the RNA can be labelled, by performing the reaction in the presence of [γ-^{32}P] GTP, instead of [α-^{32}P] UTP, since H4 mRNAs have a cap structure of the type $^{7}mGpppGp$ (80). This experiment would assume that the RNA is properly capped in vitro, a process shown to occur in other systems (143,144), but not tested in the case of human H4 histone mRNA in vitro transcription. Furthermore, finding a similar cap structure would not by itself indicate accurate initiation. A longer oligonucleotide containing the cap structure would have to be analyzed by

2-D fingerprinting to determine if it co-migrates with the in vivo capped, genuine mRNA-derived oligonucleotide. Data concerning the identification of such a cap-containing oligonucleotide by 2-D fingerprint analysis of T_2 digests of in vivo-labeled H4 mRNA have not been conclusive.

Other methods currently used to assess the specificity of initiation of in vitro transcription involve hybridization of the in vitro transcripts with a specific, labelled DNA fragment, either spanning the putative 5' end of the RNA (S_1 nuclease method), or containing only sequences complementary to the RNA transcript, but not spanning the putative 5' end (primer extension method). In the S_1 nuclease method (151,152), the protruding, unhybridized DNA sequences are removed by digestion with S_1 nuclease, which degrades single-stranded regions of DNA or RNA. The protected, shortened DNA fragment is then dissociated from its RNA complement and is sized on an appropriate gel. Its size is compared with the length of a fragment that had been hybridized to in vivo synthesized RNA, and subjected to the same S_1 nuclease digestion, denaturation and electrophoretic fractionation. The presence of protected DNA bands of similar size in both cases would indicate specific initiation in vitro (151).

In the second method, called "primer extention method" (155), a 5' end labelled DNA fragment containing complementary sequences internal to the RNA, but not spanning its putative 5' end is hybridized separately to in vitro synthesized unlabelled RNA, and to in vivo synthesized, unlabelled RNA. The hybrid molecules are then used as primers for AMV reverse transcriptase, which will elongate the DNA primer, using the

hybridized RNA as a template. The size of the elongated DNA fragment depends on the length of the RNA to which the DNA primer was hybridized (155).

These last two methods use the same principle, and the same information can be gained from either one. We chose to use the primer extension method because, being based on a synthetic step (by AMV reverse transcriptase), as opposed to a degradative step (by S_1 nuclease), the results should, in principle at least, be more reliable, and less prone to artifacts, such as those caused by secondary structure of the DNA or the RNA.

The DNA primer used for the initial experiments was a 317 bp Sac II fragment, containing nucleotides coding for the 3' end of the H4 mRNA (Figure 16A), extending from the nucleotides coding for amino acid 56 to 107 nucleotides past the 3' end of the mRNA. The DNA fragment was labeled by kinase and run on a 5% polyacrylamide gel under denaturing conditions, so as to allow strand separation. The two well resolved strands were individually isolated and used to anneal with the transcripts obtained in vitro from Eco RI-digested pFO 108. Primer extension analysis was performed separately with each DNA strand and the results were analyzed on a 6% polyacrylamide/8.3 M urea gel. Figure 21 shows that both the in vivo synthesized HeLa mRNA and the in vitro synthesized RNA hybridized to the more slowly migrating strand, and reverse transcription produced a band with exactly the same migration in both cases. Sequencing data have shown that this strand is the one that is complementary to the mRNA.

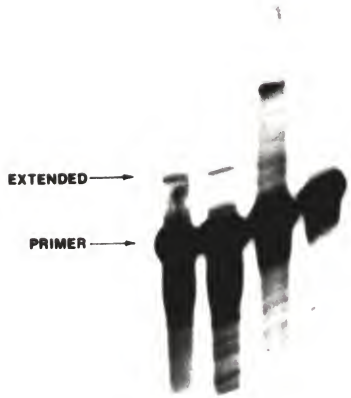
The more rapidly migrating strand hybridized to an RNA species present in the in vitro synthesized sample, but not in its in vivo counterpart. Extension of the primer after this hybridization gave rise

Figure 21: Primer extension analysis of *in vitro* transcripts from pFO 108 DNA.

Autoradiogram of a 6% polyacrylamide/8.3 M urea gel showing the size of the DNA obtained after extension of the 317 base-long primer by AMV reverse transcriptase, using HeLa polysomal RNA (lanes 1 and 3) or *in vitro* synthesized, pFO 108 DNA-directed RNA (lanes 2 and 4). In lanes 1 and 2, the RNAs were hybridized to the more slowly migrating strand of the primer. In lanes 3 and 4, the RNAs were hybridized to the more rapidly migrating strand of the primer.

The primer, as well as extended DNA primer molecules discussed in the text, are indicated by arrows.

1 2 3 4



to a DNA fragment larger than that obtained with the more slowly migrating strand. These results most likely indicate that accurate initiation of transcription at the bona fide 5' end of the H4 histone mRNA was indeed obtained *in vitro*. The results obtained with the more rapidly migrating strand probably indicate that this strand of the template is read end-to-end in the *in vitro* system.

The results obtained in this experiment are not definitive, due to the large size of the primer used, which made measurements of lengths rather inaccurate. For this reason, the experiment was repeated later, using a shorter DNA primer, defined by direct sequencing analysis. Results of these experiments will be described later.

In vitro Transcription of an H4 Gene

Having proved that specific initiation of transcription was obtained in the *in vitro* system using as a template the H4 gene present in pFO 108, several experiments were devised to determine which nucleotide sequences are required for the *in vitro* transcription of this gene.

A. The 3' Flanking Region:

The 3' flanking region of the H4 gene present in pFO 108A has been sequenced up to 107 nucleotides past the putative 3' end of the mRNA (Figure 11). The sequence shows all the features expected for a histone gene. The mRNA is most likely terminated *in vivo* at the ACCA motif found a few nucleotides downstream from the hyphenated dyad symmetry, which is characteristic of histone mRNA 3' ends (54). Interestingly, this region of the RNA can, theoretically at least, form a characteristic stem and loop structure, similar to that which appears to be involved in the termination of transcription by RNA polymerase III (165). This sequence has been found by Birchmeier *et al*. (200) to be necessary, although not

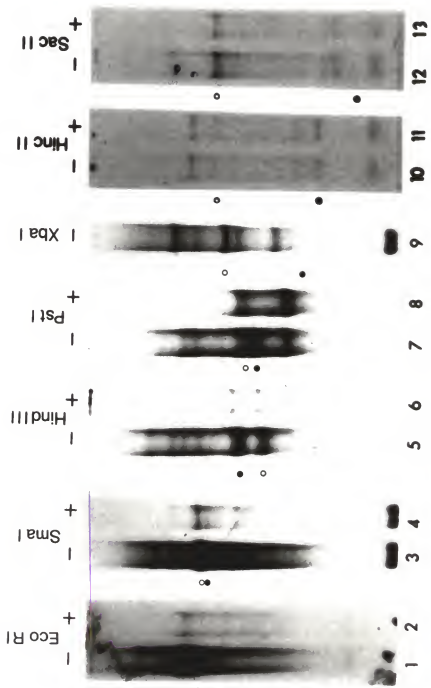
sufficient, for termination of transcription at the appropriate place, since deletion of the hyphenated dyad symmetry from a cloned sea urchin H2A gene induced the read-through by RNA polymerase II in frog oocytes. However, reinsertion of the same motif in the middle of an H2B gene did not by itself promote termination of transcription of this point. Twelve nucleotides downstream of the ACCA motif, there is another histone gene-related motif, characterized mainly by its high A+G content. No specific function has yet been ascribed to these sequences.

To determine if sequences downstream from the 3' end of the H4 histone gene are required for in vitro transcription, the initial approach was to transcribe in vitro pFO 108 DNA that had been digested with several different restriction endonucleases, shown in Figure 14A. All the experiments to be described in this section were performed using double digestions of pFO 108 DNA. DNA was first digested to completion with Eco RI restriction endonuclease, which separates the whole insert from vector sequences, thus eliminating the possibility of interference due to the presence of promoters in the pBR 322 vector. The DNA was then subjected to digestion by each one of the other restriction endonucleases indicated in the map in Figure 14A. The DNA was deproteinized by phenol extraction and transcriptions were performed in parallel in the presence and absence of 2 ug/ml of α-amanitin to determine transcripts produced by RNA polymerase II, the enzyme responsible for transcription of the histone genes in vivo (240).

Figure 22 shows an example of the type of results obtained when pFO 108 DNA digested with a series of different restriction endonucleases was transcribed in vitro, and the resulting transcripts were analyzed on a 1.5% agarose/formaldehyde gel. The black dot at the left of lanes 1, 3,

Figure 22: In vitro transcription of pFO 108 DNA digested with different restriction enzymes.

Autoradiograms of 1.5% agarose, 3% formaldehyde gels showing in vitro transcripts obtained when using as a template pFO 108 DNA that had been previously digested with Eco RI, as well as with each one of the other restriction enzymes indicated at the top of each lane, both in the presence and absence of 2 ug/ml of α-amanitin (+ and −). The black dot at the left of each lane indicate the position of the expected read-through transcript. The open circle at the left of each lane indicate the position of the α-amanitin insensitive, insert-dependent transcript discussed in the text.


Eco R1
Sma I
Hind III
Pst I
Xba I
Hinc II
Sac II
1 2 3 4 5 6 7 8 9 10 11 12 13

5, 7, 9, 10 and 12 indicate the approximate positions expected for transcripts initiated at the correct 5' terminus of the H4 gene and terminated at the point of cleavage by the restriction endonuclease used in each case. It can be observed that in vitro transcription of pFO 108 DNA that had been digested with Eco RI gives rise to several bands, including an RNA species of about 2.8 Kb, which is the only one that is sensitive to low concentrations of α-amanitin. This is the correct size for a transcript initiated at the 5' end of the H4 gene, and terminated at the distal end of the insert. All the other transcripts observed in lane 1 of Figure 22, except for the one indicated with an open circle, are derived from the pBR 322 molecule, as can be seen in lane 1 of Figure 23. Furthermore, these bands are not sensitive to α-amanitin, and thus can most likely be atributed to RNA polymerase I or RNA polymerase III activities.

In vitro transcription of pFO 108 DNA digested with Eco RI and Sma I restriction endonucleases also gave rise to several bands, including an α-amanitin sensitive transcript of about 2.7 Kb, corresponding to the expected H4 transcript. It should be emphasized that Sma I cleaves pFO 108 DNA twice, once at the 3' distal end, thus the transcript observed is shorter than when DNA digested with Eco RI is used. This enzyme also cuts pFO 108 DNA nine nucleotides from the Eco RI site located close to the 5' end of the gene. All the other bands in lane 3, except for the one indicated with an open circle, are derived from pBR 322, as shown in Figure 23, lane 1 (Sma I does not cut in pBR 322 (241), so that the Eco RI digest of pBR 322 is the appropriate control for an Eco RI/Sma I double digest of pFO 108 DNA).

In vitro transcription of pFO 108 DNA digested with Eco RI and Hind III again produced the expected results, and will not be analyzed in

Figure 23: *In vitro* transcription of pBR 322 DNA digested with different restriction enzymes.

Autoradiogram of a 1.5% agarose, 3% formaldehyde gel showing *in vitro* transcripts obtained when using as a template pBR 322 DNA that had been previously digested with Eco RI, as well as each one of the other restriction enzymes indicated at the top of each lane.


Eco RI
Pst I
Hind III
Hinc II
1
2
3
4

In vitro transcription of pFO 108 DNA digested with Eco RI and Hind III again produced the expected results, and will not be analyzed in further detail. There is again one RNA species, however, of approximately 1.0 Kb in length, which is not α-amanitin sensitive and is not derived from pBR 322. This band, indicated by an open circle in Figure 22, will be separately discussed at a later point.

When pFO 108 DNA was digested with Eco RI and Pst I restriction endonucleases, the predicted α-amanitin sensitive transcript was observed.

If pFO 108 DNA is digested with enzymes that cleave the DNA at sites closer to the 3' end of the H4 gene, such as Xba I or Hinc II, or with an enzyme that actually truncates the gene by cutting between nucleotides 165 and 166 of the H4 coding region (Sac II), no α-amanitin sensitive transcripts were observed (Figure 22). All the other bands in lanes 9, 10 and 12 in Figure 22, except for those indicated with an open circle, are derived from pBR 322 digested with the appropriate restriction enzyme (Figure 23). Notice that when the DNA is digested with Eco RI and Hinc II restriction endonucleases, a transcript of the expected size was indeed observed, however, this transcript was also obtained when pBR 322 DNA was transcribed after digestion with the same two enzymes (Figure 23, lane 4). Furthermore, this transcript is not sensitive to α-amanitin, so it was concluded that this band does not represent a true transcript of the H4 gene present in pFO 108.

An interesting transcript was observed during the course of these studies. As the template was truncated closer to the 3' end of the H4 gene, thus making the H4 transcript shorter, an α-amanitin insensitive

transcript was observed, that became larger (open circles in Figure 22). The initiation site of this transcript was determined to be close to the Sma I site that is furthest from the H4 gene, with its transcription occurring in the opposite orientation (white arrow in Figure 16A).

Experiments performed by Aleida Leza in our laboratory had indicated the presence in pFO 108 of at least one member of the Alu family of repetitive DNA sequences (202). At the same time, several investigators in our laboratory have found that, when using nick-translated pFO 108 DNA to probe RNA blots, dark backgrounds were obtained, the probe hybridizing almost throughout the whole lane of RNA. Since it has been reported that at least some members of the Alu family of DNA sequences are transcribed in vivo (235,236), it is probable that the Alu DNA sequence present in pFO 108 was responsible for the dark background. Cloned Alu DNA has been transcribed in vitro by RNA polymerase III (234), a fact that suggested that the α-amanitin insensitive transcript previously described could in fact be a product of the Alu DNA sequence present in pFO 108. A sub-clone was constructed from pFO 108 by Aleida Leza, which lacks the sequences between the Hind III sites and the Eco RI site distal to the H4 gene. This clone was called pFO 108 A (Figure 16B). When used in hybridization studies, this sub-clone did not produce the dark background observed with pFO 108, and yet, it hybridized strongly with H4 mRNA sequences.

A series of experiments similar to those shown in Figure 22 was performed using pFO 108 A DNA instead of pFO 108. The results in Figure 24 show that similar patterns were obtained, the main difference being the disappearance of the strong α-amanitin insensitive transcript previously assigned to the region of DNA that was deleted from pFO 108 while constructing pFO 108 A.

Figure 24: In vitro transcription of pFO 108 A DNA digested with different restriction enzymes.
Autoradiogram of 1.5% agarose, 3% formaldehyde gels showing in vitro transcripts obtained when using as a template pFO 108A DNA that had been previously digested with Eco RI and Hind III, as well as each one of the other restriction enzymes indicated at the top of each lane, both in the presence and absence of 2 ug/ml of α-amanitin (+ and -). The black dot at the left of each lane indicate the position of the expected read-through transcript.

I
EcoR I
+ Hind III
I
Xba I
+
I
Hinc II
+
I
Sac II
+

Taken together, results from in vitro transcription of restriction enzyme-digested pFO 108 and pFO 108 A DNAs show that, under appropriate conditions, specific RNA polymerase II transcripts as long as 2.8 Kb can be produced. However, removal of sequences downstream from the 3' end of the gene had a clear inhibitory effect on the in vitro transcription of pFO 108 or pFO 108 A DNA. Specifically, digestion of either template with Pst I restriction endonuclease, which cuts approximately 800 bp downstream from the 3' end of the gene greatly reduced the amount of specific transcripts produced under the standard conditions of the assay. Digestion of either template with enzymes that cut closer to the 3' end of the gene completely abolished the production of transcripts of the expected size. Analysis of these transcription products on 5% polyacrylamide gels containing 8.3 M urea failed to indicate the presence of smaller, specifically terminated transcripts (data not shown). These results suggested that regions located at the 3' end of the gene might have an effect in enhancing in vitro initiation and/or elongation of transcription of the H4 histone gene present in pFO 108.

To further explore these possibilities, a new series of primer extension experiments was designed, with the aim of determining if initiation of transcription had indeed occurred in cases where no specific transcript was observable on 1.5% agarose/formaldehyde gels.

By this time, sequencing data for the entire H4 coding region in pFO 108 A were available. Using this information, a 64 bp Alu I/Hha I fragment, containing those nucleotides coding for amino acids 17 through 38 (see Figure 11), was isolated from a 10% polyacrylamide gel and used as a primer to assay for specific initiation of transcription. If accurate initiation had occurred, this DNA primer would be expected to be elongated

to a molecule of approximately 160 nucleotides in length. Such a molecule could be accurately sized on a 10% polyacrylamide/8.3 M urea gel, using as molecular weight markers two sequencing ladders (A+G and C+T) derived from the 408 bp Eco RI/Sac II fragment from pFO 108 A. Under these conditions, variations in length as small as $\pm$ 2 nucleotides would be easily distinguishable.

Figure 25 shows the results of primer extension experiments performed using the 64 bp DNA fragment described above. Lane 1 shows the extended primers obtained with HeLa polysomal RNA. Two bands are clearly observable, indicating that at least some of the different H4 mRNAs (204) have enough sequence homology with the 64 bp fragment from pFO 108 A to form stable hybrids; however, differences in the 5' leader probably account for the microheterogeneity observed in the extended primer molecules.

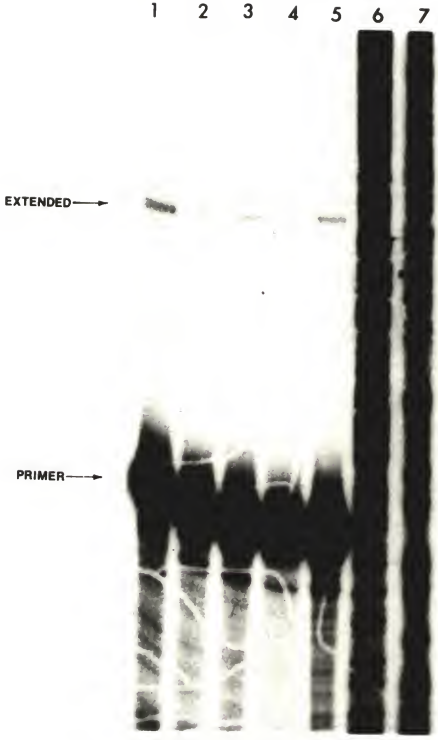
Lanes 2, 3, 4 and 5 show the extended primers obtained with RNA transcribed *in vitro*, using as a template pFO 108 A DNA digested with Eco RI plus either Hind III (lane 2), Pst I (lane 3), Xba I (lane 4) or Hinc II (lane 5). In all cases, a single band, comigrating with one of the two extended primers observed with HeLa polysomal RNA, was obtained.

These results confirmed the previous finding that specific and accurate initiation of transcription of the H4 histone gene in pFO 108 A does occur *in vitro*. Most interesting, specific initiation of transcription was observed with all the templates analyzed, in spite of the fact that with two of them (pFO 108 A DNA digested with restriction endonucleases Xba I and Hinc II), no specific, α-amanitin sensitive band was observed in 1.5% agarose/formaldehyde gels. Furthermore, as far as this semi-quantitative assay can show, the apparent rate of initiation of transcription was similar in all cases.

Figure 25: Primer extension analysis of in vitro transcripts.

Autoradiogram of a 10% polyacrylamide/8.3 M urea gel showing the DNA obtained after extension of the 64 bp primer by AMV reverse transcriptase, using different RNA samples as templates. Lane 1: HeLa polysomal RNA. Lanes 2 through 5: In vitro transcripts obtained using as a template pFO 108A DNA digested with Eco RI restriction endonuclease, as well as: Lane 2: Hind III. Lane 3: Pst I. Lane 4: XbaI. Lane 5: Hinc II. Lanes 6 and 7 are sequencing ladders obtained from the 408 bp Eco RI/Sac II fragment of pFO 108A, and used as molecular weight markers. Lane 6: A+G. Lane 7: C+T.

The positions of the unextended and the extended primers are indicated by arrows.


1
2
3
4
5
6
7
EXTENDED
PRIMER

Taken together, the results seem to indicate that truncating the pFO 108 A template closer than about 800 nucleotides from the 3' end of the H4 gene strongly inhibits elongation of transcription, since initiation still occurs at approximately normal rates and at the same site, yet no full size transcript is produced.

Although completely unexpected, these findings correlate with those reported by Grosschedl and Birnstiel (182). These authors have found that digesting a sea urchin histone H2A gene template with Eco RI, about 750 bp downstream from the 3' end of the H2A gene, produced a lower level of initiation of in vitro transcription of this gene than when the template was cut with either Hind III or Bam HI, whose recognition sites lie at positions -195 and -460 nucleotides upstream from the 5' end of the gene. Their results, however, also indicate that if the template is not cut at all, but left as a circular molecule, the same low level of initiation of in vitro transcription obtained after Eco RI linearization was observed. They concluded that there is no inhibitory effect caused by Eco RI digestion at the 3' flanking region, but rather, there is an enhancing effect on initiation of in vitro transcription promoted by digestions that produce free ends in the 5' flanking region, probably due to the creation of artificial entry sites for the RNA polymerase (182). In this respect, we have also observed a decreased level of in vitro transcription of the H4 gene present in pFO 108 A when the plasmid is not linearized with Eco RI restriction endonuclease, which produces a free end in the molecule at position -200 upstream from the 5' end of the gene. These free ends might indeed be providing additional entry sites for the RNA polymerase in the experiments described above.

B. The 5' Flanking Region:

Figures 11 and 26 show the sequences preceding the initiation codon of the H4 histone gene present in pFO 108A. Analysis of these sequences indicates the presence of several putative regulatory sequences that might be involved one way or another with the activity of this gene in vivo. About 10 bp upstream from the TATA box, there is a motif, GTTCC, very similar to the GTACC motif found in an analogous position in several sea urchin histone genes (54). Although it is well conserved, the significance of this homology block is presently unknown. Further upstream, and indicated in Figure 26 by closed boxes, there are two tandem "CAAT" boxes (181,188), remarkably similar to those found in other genes served by RNA polymerase II, including several histone genes. Usually, one of these boxes is found in the 5' flanking region of most eukaryotic genes, and in some cases, like the H2A histone gene present in the sea urchin clone h22, two of them are found in tandem arrangement (54). Interestingly, no such homology block has been found in the 5' flanking region of any H4 or H1 gene studied (54), yet the H4 gene present in pFO 108A does contain two of them, in a tandem arrangement.

Further upstream from the H4 gene, there are several other non-random sequences. Notably, between nucleotides -152 and -174 there is a stretch of 21 nucleotides which only contains A and G residues, most usually in the form of the trinucleotide GGA. These are indicated in Figure 26 by a waving underlining. Similar, although not identical stretches have been found in the spacer regions of other histone genes (54,81), and a role in recombination has been proposed, although not directly tested. Finally, several short repeats (indicated by horizontal arrows in Figure 26) are present in the 5' flanking region of the H4 gene. Although not previously

reported in the 5' flanking regions of eukaryotic genes, it is well documented that direct repeats present in the SV40 genome, and at the end of retrovirus genomes do act as strong promoters of transcription, both in vivo and in vitro (242,243).

In order to test the functional relevance of these putative regulatory sequences, a series of deletion mutants was constructed from pFO 108A, which spanned almost all the 5' flanking regions of the H4 gene, but did not include the TATA box. The TATA box has been implicated in numerous sytems as playing a role in directing the precise site of initiation of transcription by RNA polyerase II (145,154,181,182).

The clones were constructed by exonuclease digestion with BAL-31, followed by the addition of Eco RI linkers and cloning into pBR 322. After screening for the appropriate recombinants, 30 clones were characterized with respect to the size of the deletions by Eco RI/Sac II double digestion, followed by 3' end labelling and gel electrophoresis on a 3% agarose gel. Figure 26 shows the sequences upstream from the H4 gene, indicating the location of the deletion points determined for selected clones (vertical arrows).

From this collection, several clones were selected, grown and their DNA was isolated, restricted with Hind III restriction endonuclease, either alone, or together with Eco RI. Figure 27 shows the in vitro transcripts obtained from a representative sample of these clones. While it is clear that the assay can not be used in a quantitative manner, it is also obvious that all the clones under study gave rise to an in vitro transcript of the expected size, including clone pFO 108A 5'Δ80, a clone that is devoid of the direct repeats, as well as the "CAAT" boxes previously described. It should also be noticed that the same clones gave

Figure 26: Endpoints of 5' deletion mutants of pFO 108A.

Sequences upstream from the 5' end of the H4 coding region are shown. Vertical arrows indicate the end point of the different deletion mutants. Nucleotide residues are numbered from the AUG initiation codon, and decrease in the upstream direction. The deletion clones are designated after the nucleotide at which the endpoint of the deletion lies.

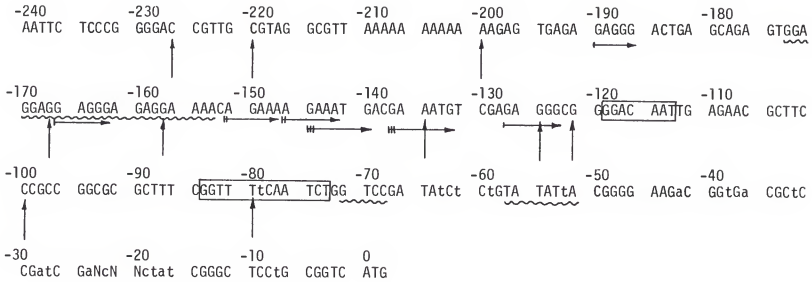
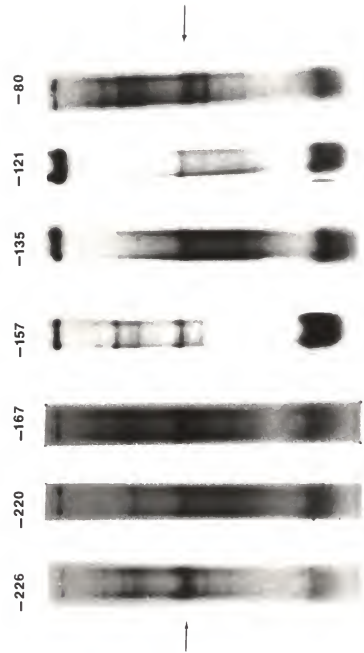

-240 -230 -220 -210 -200 -190 -180
AATTC TCCCG GGGAC CGTTG CGTAG GCGTT AAAAA AAAAA AAGAG TGAGA GAGGG ACTGA GCAGA GTGGA
-170 -160 -150 -140 -130 -120 -110
GGAGG AGGGA GAGGA AAACA GAAAA GAAAT GACGA AATGT CGAGA GGGCG GGGAC AATTG AGAAC GCTTC
-100 -90 -80 -70 -60 -50 -40
CCGCC GGCGC GCTTT CGGTT TtCAA TCTGG TCCGA TAtCt CtGTA TATtA CGGGG AAGaC GGtGa CGCtC
-30 -20 -10 0
CGatC GaNcN Nctat CGGGC TCCtG CGGTC ATG

Figure 27: In vitro transcription of deletion mutants from pFO 108A.

Autoradiograms of 1.5% agarose, 3% formaldehyde gels showing the transcripts obtained from selected deletion mutants. The template DNA was digested with Eco RI and Hind III restriction endonucleases. The arrows indicate the position of the expected read-through transcript. Numbers above the lanes refer to the position of the deletion end point with respect to the AUG initiation codon.


-80
-121
-135
-157
-167
-220
-226

a much lesser degree of in vitro transcription when the DNA template was digested only with Hind III, and not with Eco RI. Apparently, Eco RI cleavage can provide an artificial site of entry for the RNA polymerase, which then appears to start transcribing at its normal site.

Even though the in vitro transcripts observed with all of the deletion mutants did not strongly differ from those obtained when using the parental plasmid, pFO 108A, as a template, further evidence for the involvement of RNA polymerase II in the production of in vitro transcripts from templates containing deletions was assessed by performing similar reactions in the presence and in the absence of 4 ug/ml of α-amanitin. Figure 28 shows that, as previously observed for the parental plasmid, the synthesis of a 1.6 Kb RNA transcript by the in vitro transcription system is inhibited by low concentrations of α-amanitin, a result which implies the participation of RNA polymerase II in the in vitro synthesis of this transcript.

Primer extension experiments performed as previously described indicate that the in vitro transcript observed when using clone pFO 108A 5'Δ80 as template is initiated at the bona fide 5' end initiation site, as compared with HeLa cell polysomal RNA (Figure 29).

The results then clearly indicate that no sequences upstream from the TATA box are required for the in vitro transcription of the H4 gene present in pFO 108A.

Figure 28: In vitro transcription of deletion mutants from pFO 108A, both in the presence and absence of 4 ug/ml of α-amanitin.

All the DNA templates were digested with Eco RI and Hind III restriction endonucleases. Lanes 1 and 2: pFO 108A. Lanes 3 and 4: pFO 108 A 5'Δ 225. Lanes 5 and 6: pFO 108A 5'Δ 167. Lanes 7 and 8: pFO 108A 5'Δ 80. + and - signs on top of each lane indicate the presence (+) or absence (-) of 4 ug/ml of α -amanitin in the transcription reaction.

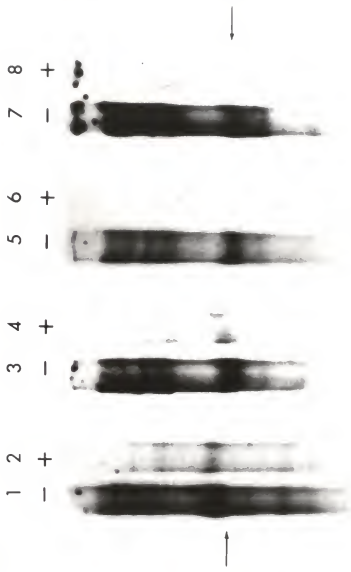
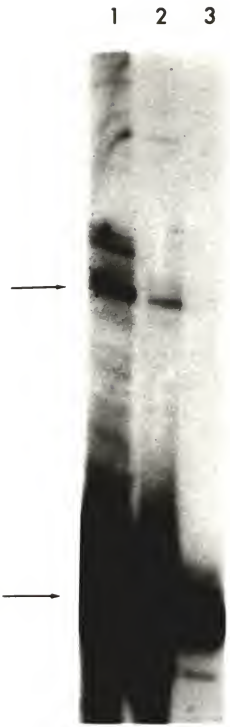

1 2
− +
3 4
− +
5 6
− +
7 8
− +

Figure 29: Primer extension analysis of transcripts generated using pFO 108A 5'Δ80 as a template.

Autoradiogram of a 10% polyacrylamide/8.3 M urea gel showing the DNA obtained after extension of the 64 bp primer by AMV reverse transcriptase, using different RNA samples as templates. Lane 1: HeLa polysomal RNA. Lane 2: *in vitro* transcripts from pFO 108A 5'Δ80. Lane 3: Unhybridized primer. The positions of the unextended and the extended primers are indicated by arrows.


1
2
3

DISCUSSION

Histone genes from several species have been cloned and analyzed in recent years. The genes expressed during the early development of sea urchins were the first histone genes to be studied in detail (83-88), followed by those of the fruit fly Drosophila melanogaster (102). Those genes proved to be clustered, and tandemly repeated in simple units, containing one each of the five histone genes, H1, H2A, H2B, H3 and H4. The genes are separated from each others by short, A+T rich spacer regions. In sea urchins, all the genes are arranged with the same polarity, while in Drosophila melonogaster, different genes are read from different strands (81).

In the case of yeast, 2 copies of each histone gene have been found per haploid genome. Cloning of these DNA fragments has shown that the H2A and H2B genes are adjacent to each other, but their transcriptional polarity is divergent. Furthermore, the H3 and H4 genes are not found adjacent to the H2A and H2B genes (105). More recent research has shown that the organization of histone genes in vertebrates is far more complicated than shown for the aforementioned species. In the case of Xenopus laevis, the genes still appear to be clustered, and tandemly repeated, however, more than one gene order has been found, each one associated with a different H1 variant (111,112). In the case of the newt Notophthalamus viridescens, homogeneous 9 Kb clusters, each containing one of each of the five histone genes have been found; however, these clusters are not arranged in tandem repeats, but they are independent and separated by up to 50 Kb of non-related DNA sequences (106,109). In the case of

chicken, mouse and humans, including the human histone genes described here, histone genes have been found to be organized in clusters, however, no simple tandem repeats have been observed from the type of analysis reported to date.

Studies on the genomic organization of human histone genes presently being conducted in our laboratory by Rik van Antwerpen, using the λHHG clones as sources of probes, have indicated the likelihood that clones λ HHG5, λHHG41 and λHHG55 are independent representatives from a rather abundant type of repeat present in the human genome in several copies. These results are however not conclusive yet, since only probes derived from these same clones have been used, which raises the possibility that the higher degree of hybridization observed in the genomic blots might be due to a higher degree of homology between the specific probe being used and the genomic sequence from which this probe was isolated. Other variables that should be considered while analyzing genomic blots include (1) the length of the probe, since DNA regions adjacent to a specific gene will hybridize with themselves if those DNA sequences are present in the probe, thus making the hybridization signal of this particular gene much stronger than that observed for other genes that do not share the same flanking regions with the probe; (2) the physical size of the DNA fragments present in the probe after nick-translation. High specific activity probes are required for adequate analysis of genomic blots, however, at very high specific activities, the DNA is usually sheared to very small fragments, that will further emphasize problems like those discussed under 1.

Other investigators have cloned human histone genes and of those isolates reported in the literature (118,123), none shows analogous restriction maps with clones of the λHHG 41 group, while some of the

clones reported by Heintz et al. (123) do correspond approximately to the restriction map of clones of the λHHG 17 group and maybe to the restriction map of clone λHHG 39. These facts seem to argue against the notion that λHHG 41 is a representative of a major repeat of human histone genes. However, this interpretation might be biased, since Heintz et al. used an H4 cDNA probe to select their clones, while I used a probe containing chicken H3 plus H4 genes. Consequently, the selection of strongly positive hybridizing phage plaques does bias the selection of genomic clones towards clones containing multiple copies of the H4 histone gene. In the case of the clones isolated by Clark (118), a probe containing all four core histone genes from chicken (H2A, H2B, H3 and H4) was used to select two human histone gene clones. Neither clone contains H4 coding sequences, and the patterns of histone gene organization do not overlap with those described in this dissertation or those described by Heintz et al. (123). For these reasons, the clones that have been reported might not be a statistically significant sample of the whole range of possible arrangements of the histone genes in the human genome.

Hybridization data, however, indicate that the H4 and H3 genes present in λHHG 17 hybridize much less efficiently with HeLa total RNA than do the genes from λHHG 41, a fact that again argues that λHHG 41 might indeed be a representative of a more repeated type of cluster. An alternative explanation is that the genes in the λHHG 41 group have promoters that are stronger than those found in the genes from the λHHG 17 group, thus making their transcripts more abundant in HeLa cells, as has been shown by Lichtler et al. (204).

It is theoretically possible that all of the genes present in clones of the λHHG 41 group are coordinately expressed in HeLa cells to a higher extent than their counterparts in the cluster represented by λHHG 17. In

that respect, it is interesting to note that, so far, two 5' flanking regions of genes present in λHHG clones have been sequenced: that of the H4 gene present in λHHG 41, presented in this dissertation, and that of the H2B gene present in λHHG 55 (Dr. Farhad Marashi, personal communication). From these two, experiments performed together with Dr. Alex Lichtler have shown that the H4 gene present in λHHG 41 is completely complementary over its entire length with one of the major species of H4 mRNA found in HeLa cells (204). This gene shows a prelude region containing putative regulatory sequences, including the A+G-rich region at position -200, two "CAAT" boxes, and a TATA box. Furthermore, the coding region has the capacity to encode an H4 histone protein identical to that found in calf thymus, and the 3' end of the mRNA shows the usual T-hyphenated dyad symmetry found in most histone genes, followed by the ACCA termination motif (54). In other words, this gene seems to be completely functional, as far as the available assays can determine. On the other hand, the H2B gene sequenced by Dr. Farhad Marashi probably does not code for an H2B protein, since it contains two frame-shift mutations, as well as several amino acid substitutions, some of which correspond to well conserved, functionally important residues, such as the Pro→Val substitution at position 103 (101). Furthermore, no TATA box has been detected within 132 nucleotides upstream from the AUG initiation codon, and no termination codon exists at the appropriate position. The mutations present in the H2B gene from λHHG 55 could conceivably have been introduced during the cloning or the subcloning of the gene; however, it is reasonable to expect that these sequences are the ones found in vivo. This result could be interpreted in two possible ways: (1) In the same gene cluster, at least one actively transcribed gene might co-exist with a pseudo-gene. This type of arrangement has been described as existing both

in the α and the β-globin gene clusters, where pseudogenes are interspersed with active members of each of these gene families (247,248). (2) It is possible that a whole cluster (the one present in λHHG 55) is aberrant, while a similar, though not identical (see results) cluster (the one present in λHHG 41) might be fully active. Sequencing of the H4 gene present in λHHG 55 would most certainly distinguish between these possibilities.

Three different laboratories have reported the isolation and characterization of human histone genes (118,122,123). Considering that clones λHh 2 and λHh 7, isolated by Heintz et al. (123) share a similar restriction map, as well as histone gene organization map, with clones λ HHG 39 and λHHG 17, respectively, we can conclude that 8 different clusters have been isolated. In all of these clusters there are a total of 9 different H4 coding regions. Since the histone genes are reiterated 20-40 times per human haploid genome (227), it is still possible that other arrangements containing H4 histone genes have not yet been cloned. Alternatively, if any type of repeat does indeed exist, it is possible that all or most of the H4 histone genes present in the human genome have already been cloned and isolated. In this respect, it is interesting to note that Lichtler et al. have found between 7 and 8 different H4 mRNAs in HeLa cells (204). This number is close to the total number of histone genes that have been cloned so far.

Different H4 mRNAs in HeLa cells, which have been shown by tryptic mapping to code for the same H4 protein (80), may differ from each other in primary structure in the non-translated leader and trailer regions, as well as in iso-coding codons (wobbling). Comparisons including only coding regions for H4 mRNA have shown the following degree of divergence, when comparing the DNA coding regions of different species with the DNA coding regions of the human histone gene present in pFO 108A:

	% Nucleotide Divergence	% Aminoacid Divergence
Sea urchin (S. purpuratus)	22.2	0
Xenopus laevis	17.3	0
Mouse	12.7	0

This analysis indicates that, whenever the evolutionary span allows, mutations have occurred as much as it is possible without changing the coding capacity of the H4 mRNA in question. If we postulate that the same is true for histone genes within the human genome, we should expect that the 20-40 copies of histone genes in the human genome would give rise to 20-40 different H4 mRNA. This postulate is realistic if we consider that the two H2B genes present in yeast differ from each other by 12.6% of the nucleotides present in the protein coding regions (105). Similarly, two H4 genes present in different early repeats of the histone genes from the sea urchin Psammechinus miliaris show as much as 10.3% divergence in the nucleotide sequence of the coding region (93,98). On the other hand, it is possible that not all of these 20-40 copies of each histone gene are able to produce a functional mRNA. Such is the case for the H2B gene present in λHHG 55. This gene has been sequenced by Dr. Farhad Marashi, and the mutations observed in its sequence indicate that it can not code for a functional H2B protein.

This analysis does not take into consideration the possibility of variations in the leader or trailer regions; yet it still indicates that (1) it is not surprising to find several different mRNAs coding for the same proteins, when those mRNAs originate from a family of middle repetitive genes, such as the histone genes, and (2) actually, more than 8 or 9 mRNA species coding for an H4 protein should be detectable if a more powerful resolution technique were available.

The isolation and characterization of clones containing human histone genes described in this dissertation has given us a tool for approaching several biological problems concerning histone gene expression. I have already discussed what we have learned about the organization of the human histone genes. Work is currently in progress for the elucidation of the gross arrangement of human histone genes in several cell lines, by using genomic Southern blot analysis.

A more interesting question concerns the level of regulation of histone gene expression. In HeLa cells (65-71), as well as in yeast (249), it has been shown that histone genes are preferentially, though not solely, transcribed during the S phase of the cell cycle. This temporal, differential gene activity seems to be regulated, as least in part, at the transcriptional level. Dr. Mark Plumb in our laboratory is using the subclones derived from λHHG phage for hybrid selection of newly synthesized, in vivo labelled histone mRNA from HeLa cells. Comparison of the results obtained in those experiments with the results obtained when total accumulation of histone mRNA is measured by northern blot analysis, again using the subclones from the λHHG phage as probes, has shown that while the relative abundance of histone mRNA during the S phase of the cell cycle seems to parallel the relative rate of DNA synthesis, the synthesis of histone mRNA occurs as a burst shortly after the onset of DNA replication (75). The possible coupling between the triggering of DNA replication and/or histone gene replication with the triggering of histone mRNA biosynthesis is currently being studied.

These, as well as numerous previous studies, have shown that the accumulation of histone mRNA in early S phase of HeLa cells depends at least partially on the rates of transcription of the genes (65-71). This

indicates the necessity for the existence of some type of control mechanism, whereby it would be possible to turn the gene's transcription on or off, according to the physiological requirements of the cell. In vivo studies only indicate what effect different stimuli might have on the transcription rates of the histone genes, as well as to what extent different genes are being transcribed at any given point, since it is conceivable that different variants might be preferentially expressed at different times during the cell cycle (73) It is possible that some genes are expressed in certain cell types, but not in others or, as in the case of sea urchins, some genes might only be expressed at some developmental stages. Elucidation of the molecular mechanisms involved in the regulation of histone gene activity will necessarily require both the information obtained from in vivo studies and from in vitro studies, where transcription is studied in less physiological, although more controlled conditions.

An in vitro transcription system capable of supporting the transcription of a human histone gene, as well as its use in delineating those DNA sequences that are required for in vitro transcription of an H4 gene are presented. The system chosen for these studies is the whole cell extract from HeLa cells described by Manley et al. (144). Initial experiments, using truncated templates derived from pST 519 (H3) or from pFO 108 (H4) were unsuccessful, and it eventually became apparent that the standard procedure of truncating the template at a position close to the site of initiation of transcription was not functional in the case of these human histone genes. On the contrary, a very long stretch of DNA further downstream from the 3' end of the H4 mRNA from pFO 108 is still required for specific transcription of this gene in vitro.

By using the complete 3.1 Kb insert from clone pFO 108 as a template, it was possible to standardize the system with respect to template concentration, lysate concentration and nucleotide triphosphate concentration, as well as the effect of a 15 minute chasing period in the presence of 1 mM UTP and the effectiveness of different isolation procedures for the analysis of in vitro transcripts. After standardization with respect to all of these parameters, the results obtained in different experiments were qualitatively reproducible; however, quantitation of levels of transcription was not possible, due to a large degree of variability observed from experiment to experiment. These results most probably reflect a nigh sensitivity of the in vitro transcription system to impurities present in the DNA. Alternatively, the lack of a reproducibly quantitative recovery of in vitro transcripts might be a reflection of the large size of the transcripts obtained from pFO 108 DNA, as compared with those obtained by other investigators, using truncated templates. In their hands, the in vitro transcription system seems to be suitable for quantitative analysis (145).

Several genes, such as many adenovirus genes (143,144,151,152), α- and β-globin (153-155), ovalbumin (152) and conalbumin (141,152) have been transcribed in vitro using the Manley system. Many of these genes, all of which are served by RNA polymerase II, transcribe in vitro in the Manley system with widely varying efficiencies. More interesting, both deletion analysis and restriction endonucleolytic cleavage of the DNA templates, have shown that, for most genes studied, including a sea urchin H2A histone gene (145), sequences upstream from the TATA box do not seem to be essential for in vitro transcription, although in some cases, quantitative differences have been reported upon deletion of sequences as far as about

100 bp upstream from the 5' initiation site (145). These differences are only quantitative and do not abolish transcription.

Qualitative differences in the in vitro transcripts obtained have been observed upon removal of the TATA box preceding several different genes. In most cases described, removal of the TATA box leads to a heterogeneous population of 5' ends in the mRNA, attributable to a lessening of the stringency of initiation specificity (145,154,181,182). On the other hand, a TATA to TAGA point mutation in the upstream region of a sea urchin H2A histone gene caused a reduction, but not a complete suppression of the H2A transcripts synthesized in vitro. The transcripts obtained also showed a heterogeneous population of 5' ends (145).

H1 and H4 histone genes reported to date lack the characteristic "CAAT" box found upstream from the TATA box in many genes served by RNA polymerase II, including other histone genes (H2A, H2B and H3) (54). Again, deletion of these sequences in the conalbumin gene (250) or the β-globin gene (181) does not prevent specific in vitro transcription, and in the case of the sea urchin H2A gene, transcription in a Xenopus oocyte system may be enhanced upon removal of the "CAAT" box. (251).

Unlike the H4 genes of sea urchins, Drosophila melanogaster and mouse, pFO 108 contains a CAAT box in its 5' flanking region. Furthermore, a modified CAAT box is also present further upstream.

In an attempt to examine further the 5' upstream sequence requirements for the in vitro transcription of this gene, a series of 5' deletion mutants were tested in the in vitro transcription system. The largest deletion obtained (pFO 108A 5'Δ80) still contained the TATA box, however, both of the "CAAT" boxes have been deleted. The results indicate that all of the deletion mutants studied were able to sustain in vitro

transcription of the H4 histone gene in pFO 108, thus confirming what has been found in other systems: no sequences upstream from the TATA box, including the CAT box, are required for the in vitro transcription of genes by RNA polymerase II in the HeLa cell extracts (143-145,151-155). These results conflict with what has been found in vivo using COS cells transfected with pSVOd derived plasmids. Preliminary results obtained in collaboration with Dr. Saul Silverstein (Columbia University) have shown that the sequences present in pFO 108 are not sufficient to sustain detectable levels of transcription in this in vivo system. However, this clone does contain enough information as to be transcribed in vitro in the HeLa cell extract described by Manley. The same type of difference between in vivo and in vitro experiments has been found for the α-globin type of genes (142). In this case, the "CAAT" box has been found to be required for in vivo expression but not for in vitro transcription.

The efficiency of in vivo transcription in the pSVOd/COS cell system between the α- and the β-globin genes varies by more than a factor of 100 (142), a fact that suggests that the lack of transcription of the H4 gene from pFO 108, when subcloned into pSVOd, does not necessarily mean that additional specific 5' upstream sequences are required. It could also mean that the H4 gene promoter is not very efficiently used in the pSVOd/COS cell system. Alternative explanations also include that the H4 gene is initially read properly in the pSVOd/COS cell system, thus giving rise to a functional histone H4 mRNA. The level of mRNA within the cell might be under feedback control, as suggested by Plumb et al. (75). Alternatively, it is conceivable that the histone mRNA is translated into large amounts of H4 protein within the COS cells. We presently know very little about the metabolism and especially the turnover rate of histone proteins

produced out of the S phase of the cell cycle, and uncoupled from DNA replication. It is reasonable to expect that these proteins and/or their mRNA could be rapidly degraded. It is also possible that the protein is not degraded, but travels to the nucleus, where it might have deleterious effects on the transcription and/or replication machineries of the COS cell, thus inhibiting the further production of H4 mRNA. Finally, the excess H4 protein might not travel to the nucleus, but remain in the cytoplasm, where it could associate itself with plasmid DNA (the pSVOd molecule containing the H4 histone gene insert), thus inhibiting the transcriptional capacity of the H4 histone gene.

An unexpected involvement of the 3' downstream regions of pFO 108A with the in vitro transcription system has been observed throughout the transcription studies. A more detailed analysis of the DNA region at the 3' end of the gene that was required for in vitro transcription was done by using several restriction endonucleases to truncate the template at different positions. As was observed in the preliminary studies, truncating the template with Sac II, which cuts the template at the codon coding for amino acid 56 in the H4 protein, gives rise to a template that does not support run-off in vitro transcription, as detected by the lack of an RNA band of the appropriate size on a 1.5% agarose, 3% formaldehyde gel. Furthermore, cutting the template with restriction endonucleases Hinc II or Xba I, which cut pFO 108 DNA 100 and 250 nucleotides past the 3' end of the mRNA, respectively, also renders the template unable to support run-off in vitro transcription as detected in this gel system. It is not until the template is truncated with Pst I, 800 bp past the 3' end of the mRNA, that run-off in vitro transcription of the H4 gene is again observed in formaldehyde gels.

When initiation of transcription in vitro was studied by primer extension analysis, I found that all of the pFO 108 A DNA templates are able to support initiation of transcription at the true 5' end of the H4 mRNA, regardless of the place where the gene had been truncated. This result indicates that digestion of the template with Hinc II or Xba I restriction endonucleases does not inhibit transcription initiation, even though no run-off transcript was observed.

It is conceivable that when the template is digested with a restriction enzyme that cuts close to the 3' end of the gene, such as Xba I or Hinc II, the HeLa system recognizes the termination signals in the template, and does not produce a run-off transcript as expected, but it could rather produce a correctly terminated, mRNA-sized transcript. This possibility was tested (not shown) by electrophoresing such transcripts on a tighter polyacrylamide gel. No mRNA-size transcripts were observed, and actually, no transcript was observed that differed at all from those obtained by in vitro transcription of pBR 322 digested with the appropriate restriction enzymes. Further studies of the possible involvement of sequences 3' downstream of the H4 gene on in vitro transcription will be required to elucidate this unique observation.

Taken together, the results presented in this dissertation indicate that in humans, histone genes are clustered, but no obvious repeats were observed. Cloning of human histone genes has provided our laboratory with a powerful tool for studying the mechanism(s) operating in the regulation of human histone gene expression under different biological circumstances. A subclone containing a human H4 histone gene has been used to standardize an in vitro transcription system capable of transcribing this H4 histone gene. The system has been used to define

those nucleotide sequences which are required for the accurate initiation of transcription of this gene in vitro. The results suggest that the "CAAT" boxes found at the 5' flanking region of the H4 gene are not required for accurate and specific initiation of transcription in vitro. However, sequences located downstream of the H4 gene were found to be required for accurate production of an in vitro run-off transcript. It is not yet clear if specific sequences located in this region are required, or whether the presence of any DNA, regardless of the nucleotide sequence, would suffice this 3' downstream requirement.

REFERENCES

1. Gurdon, J.A., De Robertis, E.M. and Partington, G., Nature 260, 116 (1976).

2. Marks, P.A. and Rifkind, R.A., Ann. Rev. Biochem. 47, 419 (1978).

3. Davis, M.M., Kim, S.K. and Hood, L.E., Science 209, 1360 (1980).

4. Brown, D.D. and Dawid, I.B., Science 160, 272 (1968).

5. Britten, R. and Davidson, E., Science 165, 349 (1969).

6. Britten, R. and Kohne, D.E., Science 161, 3841 (1968).

7. McKnight, G.S. and Palmiter, R.D., J. Biol. Chem. 254, 9050 (1979).

8. Perrin, F., Cochet, M., Gerlinger, P., Cami, B., LePennec, J.P. and Chambon, P., Nuc. Acids Res. 6, 2731 (1979).

9. Stein, G., Stein, J., Thrall, C. and Park, W. in "Chromosomal Proteins and their Role in the Regulation of Gene Expression", eds. Stein, G.S. and Klein-Smith, L.J. (Academic Press, New York), pp. 1-18 (1975).

10. Spalding, J., Kajiwara, K. and Mueller, G., Proc. Nat. Acad. Sci. 56, 1535 (1966).

11. Chambon, P., Cold Spring Harbor Symp. Quant. Biol. 42, 1209 (1978).

12. Revel, M. and Groner, Y., Ann. Rev. Biochem. 47, 1079 (1978).

13. Molloy, G. and Puckett, L., Prog. Biophys. Molec. Biol. 31, 1 (1976).

14. Abelson, J., Ann. Rev. Biochem. 48, 1035 (1979).

15. Shafritz, D.A. in "Molecular Mechanisms of Protein Synthesis", eds. Weissbach, H. and Petska, S., pp. 555-601 (Academic Press, New York) (1977).

16. Civelli, O., Vincent, A., Buri, J.F. and Scherrer, K., FEBS lett. 72, 71 (1976).

17. Gabrielli, F. and Baglioni, C., Nature 269, 529 (1977).

18. Blobel, G., Doberstein, B., J. Cell Biol. 67, 835 (1975).

19. Katz, T., Rothman, J., Lingappa, V., Blobel, G. and Lodish, H., Proc. Nat. Acad. Sci. 74, 3278 (1977).

20. Harmey, M.A., Hallermayer, G., Korb, H. and Neupert, W., Eur. J. Biochem. 81, 533 (1977).

21. Losick, R. and Chamberlin, M. (eds.) "RNA polymerase" (Cold Spring Harbor Laboratory, New York) (1976).

22. Jacob, F. and Monod, J., J. Mol. Biol. 3, 318 (1961).

23. Winkler, M.E., Roth, D.J. and Hartman, P.E., J. Bacteriol. 133, 830 (1978).

24. Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C.L., Squires, C. and Yanofsky, C., Science 1989, 22 (1975).

25. Derman, E., Proc. Nat. Acad. Sci. 78, 5425 (1981).

26. Triadon, P., Lelong, J.C., Gros, F. and Crepin, M., Biochem. Biophys. Res. Comm. 101, 45 (1981).

27. Nevins, J.R. and Winkler, J.J, Proc. Nat. Acad. Sci. 77, 1893 (1980).

28. Chow, L.T. and Broker, T.R., Cell 15, 497 (1978).

29. Shatkin, A.J., Cell 9, 645 (1976).

30. Perry, R.P., Ann. Rev. Biochem. 45, 605 (1976).

31. Schwartz, H. and Darnell, J.E., J. Mol. Biol. 104, 833 (1976).

32. Bawerman, G., Ann. Rev. Biochem. 43, 621 (1974).

33. Kumar, A. and Pederson, T., J. Mol. Biol. 96, 353 (1975).

34. Adesnik, M., Salditt, M., Thomas, W. and Darnell, J.E., J. Mol. Biol. 71, 21 (1972).

35. Iskhahov, B.K., Filimonov. M.G. and Ajtkhozhin, M.A., Biochem. Biophys. Acta 521, 470 (1978).

36. Suria, D. and Liew, C.C., Can. J. Biochem. 57, 32 (1979).

37. Banerjee, A.K., Microbiol. Rev. 44, 175 (1980).

38. Weintraub, H, Larsen, A. and Groudine, M., Cell 24, 333 (1981).

39. Groudine, M. and Weintraub, H., Cell 24, 393 (1981).

40. Berget, S.M., Moore, C. and Sharp, P.A., Proc. Nat. Acad. Sci. 74, 3171 (1977).

41. Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J., Cell 12, 1 (1977).

42. Schwartz, H. and Darnell, J.E., J. Mol. Biol. 104, 833 (1976).

43. Mendecki, J., Lee, S.Y. and Brawerman, G., Biochemistry 11, 792 (1972).

44. Bergman, J.E. and Brawerman, G., Biochemistry 16, 259 (1977).

45. Scott, M.P. and Pardue, M.L., Proc. Nat. Acad. Sci. 78, 3353 (1981).

46. Getz, M.J., Birnie, G.D., Young, B.D., MacPhail, E. and Paul, J., Cell 4, 121 (1975).

47. Furuichi, Y., Muthukrishnan, S. and Shatkin, A.J., Proc. Nat. Acad. Sci. 72, 742 (1975).

48. Shimotohno, K., Kodama, Y., Hashimoto, J. and Miura, K.J., Proc. Nat. Acad. Sci. 74, 2734 (1977).

49. Furuichi, Y., LaFiandra, A. and Shatkin, A.J., Nature 266, 235 (1977).

50. Lockard, R.E. and Lane, C., Nuc. Acids Res. 5, 3237 (1978).

51. Klessig, D.F., Cell 12, 9 (1977).

52. Rogers, J. and Wall, R., Proc. Nat. Acad. Sci. 77, 1877 (1980).

53. Oshima, Y., Itoh, M., Okada, N. and Miyata, J., Proc. Nat. Acad. Sci. 78, 4471 (1981).

54. Hentschel, C.C. and Birnstiel, M.L., Cell 25, 301 (1981).

55. Nagata, S., Mantei, N. and Weissmann, C., Nature 287, 401 (1980).

56. Truman, D.E.S., "The Biochemistry of Cytodifferentiation". (Blackwell Sci. Publ., London, Great Britain) (1974).

57. Ochoa, S. and deHaro, C., Ann. Rev. Biochem. 48, 549, (1979).

58. Hershko, A. and Fry, M., Ann. Rev. Biochem. 44, 775 (1975).

59. Rubin, C.S. and Rosen O.M., Ann. Rev. Biochem. 44, 831 (1975).

60. Kanungo, M.S. and Thakur, M.K., Biochem. Biophys. Res. Comm. 86, 14 (1979).

61. Paik, W.K. and Kim, S., Science 174, 114 (1971).

62. Hayaishi, O. and Ueda, K., Ann. Rev. Biochem. 46, 95 (1977).

63. Monod, J., Wyman, J. and Changeux, J.P., J. Mol. Biol. 12, 88 (1965).

64. Koshland, D.E., Jr., Nemethy, G. and Filmer, D., Biochemistry 5, 365 (1966).

65. Borun, T., Scharf, D. and Robbins, E., Proc. Natl. Acad. Sci. 58, 1977 (1967).

66. Galwitz, D. and Mueller, G., J. Biol. Chem. 244, 5947 (1969).

67. Jacobs-Lorena, M., Baglioni, C. and Borun, T., Proc. Natl. Acad. Sci. 69, 2095 (1972).

68. Marashi, F., Baumbach, L., Rickles, R., Sierra, F., Stein, J.L. and Stein, G.S., Science 215, 683 (1982).

69. Spalding, J., Kajiwara, K. and Mueller, G.C., Proc. Nat. Acad. Sci. 56, 1535 (1966).

70. Robbins, E. and Borun, T.W., Proc. Nat. Acad. Sci. 57, 409 (1967).

71. Moll, R. and Wintersberger, E., Proc. Nat. Acad. Sci. 73, 1863 (1976).

72. Tarnowka, M.A., Baglioni, C. and Basilico, C., Cell 15, 163 (1978).

73. Wu, R.S. and Bonner, W.M., Cell 27, 321 (1981).

74. Groppi, V.E. and Coffino, P., Cell 21, 95 (1980).

75. Plumb, M., Stein, G.S. and Stein, J.L., Nuc. Acids Res. (in press) (1983).

76. Staynov, D.Z., Spencer, M., Allan, J. and Gould, H.J., Nature 279, 263 (1979).

77. Elgin, S.C.R. and Weintraub, H., Ann. Rev. Biochem. 44, 725 (1975).

78. Zweidler, A. and Cohen, L.H., J. Cell. Biol. 59, 378a (1973).

89. Newrock, K.M., Cohen, L.H., Hendricks, M.B., Donnelly, R.J. and Weinberg, E.S., Cell 14, 327 (1978).

80. Lichtler, A.C., Detke, S., Phillips, J.R., Stein, G.S. and Stein, J.L., Proc. Nat. Acad. Sci. 77, 1942 (1980).

81. Kedes, L.H., Ann. Rev. Biochem. 48, 837 (1979).

82. Hackett, P.B. and Traub, P., J. Mol. Biol. 126, 619 (1978).

83. Kedes, L.H. and Birnstiel, M.L., Nature New Biol. 230, 165 (1971).

84. Birnstiel, M., Telford, J., Weinberg, E. and Stafford, D., Proc. Nat. Acad. Sci. 71, 2900 (1974).

85. Kedes, L.H., Chang, A.C.Y., Housmann, D. and Cohen, S.N., Nature 255, 533 (1975).

86. Cohn, R.H., Lowry, J.C. and Kedes, L.H., Cell 9, 147 (1976).

87. Clarkson, S.G., Smith, H.O., Schaffner, W., Gross, K. and Birnstiel, M.L., Nuc. Acids Res. 3, 2617 (1976).

88. Overton, G.C. and Weinberg, E.S., Cell 14, 247 (1978).

89. Schaffner, W., Gross, K., Telford, J. and Birnstiel, M., Cell 8, 471 (1976).

90. Kedes, L.H., Cell 8, 321 (1976).

91. Gross, K., Schaffner, W., Telford, J. and Birnstiel, M., Cell 8, 479 (1976).

92. Wu, M., Holmes, D.S., Davidson, N., Cohn, R.H. and Kedes, L.H., Cell 9, 163 (1976).

93. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L., Cell 14, 655 (1978).

94. Sures, I., Lowry, J. and Kedes, L.H., Cell 15, 1033 (1978).

95. Engel, J.D., Sugarman, B.J. and Dodgson, J.B., Nature 297, 434 (1982).

96. Cohn, R.H. and Kedes, .L.H., Cell 18, 843 (1979).

97. Cohn, R.H. and Kedes, L.H., Cell 18, 855 (1979).

98. Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L., Nuc. Acids Res. 8, 957 (1980).

99. Childs, G., Maxson, R., Cohn, R.H. and Kedes, L., Cell 23, 651 (1981).

100. Finnegan, D.J., Rubin, G.M., Young, M.W. and Hogness, D.S., Cold Spring Harbor Symp. Quant. Biol. 42, 1053 (1977).

101. Isenberg, I., Ann. Rev. Biochem. 48, 159 (1979).

102. Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S., Cold Spring Harbor Symp. Quant. Biol. 42, 1047 (1977).

103. Hereford, L., Fahrner, K., Woolford, J., Rosbash, M. and Kaback, D.B., Cell 18, 1261 (1979).

104. Hartwell, L.H., Bacteriol. Rev. 38, 164 (1974).

105. Wallis, J.W., Hereford, L. and Grunstein, M., Cell 22, 799 (1980).

106. Stephenson, E.C., Erba, H.P., and Gall, J.G., Cell 24, 639 (1981).

107. Edstrom, J.E. and Gall, J.G., J. Cell Biol. 19, 279 (1963).

108. Kuderman, J.V. and Pardue, M.L., Dev. Biol. 60, 48 (1977).

109. Gall, J.G., Stephenson, E.C., Erba, H.P., Diaz, M.O. and Barsacchi-Pilone, G., Chromosoma 84, 159 (1981).

110. Jacob, E., Malacinski, G. and Birnstiel, M.L., Eur. J. Biochem. 69, 45 (1976).

111. Moorman, A.F.M., deLaaf, R.T.M. and Destree, O.H.J., Gene 10, 185 (1980).

112. Zernik, M., Heintz, N., Boime, J. and Roeder, R.G., Cell 22, 807 (1980).

113. van Dongen, W., de Laaf, L., Zaal, R., Moorman, A. and Destree, O., Nuc. Acids Res. 9, 2297 (1981).

114. Harvey, R.P. and Wells, J.R.E., Nuc. Acids Res. 7, 1787 (1979).

115. Harvey, R.P., Krieg, P.O., Robins, A.J., Coles, L.S. and Wells, J.R.E., Nature 294, 49 (1981).

116. Engel, J.D. and Dodgson, J.B., Proc. Nat. Acd. Sci. 78, 2856 (1981).

117. Krieg, P.A., Robins, A.J., Gait, M.I., Titmas, R.C. and Wells, J.R.E., Nuc. Acids Res. 10, 1495 (1982).

118. Clark, S., Ph. D. Thesis, University of Adelaide, Australia.

119. Scott, A.C. and Wells, J.R.E., Nature 259, 635 (1976).

120. Seyler-Tuyns, A. and Birnstiel, M.L., J. Mol. Biol. 151, 607 (1981).

121. Sittman, D.B., Chiu, I.M., Pan, C.J., Cohn, R.H., Kedes, L.H. and Marzluff, W.F., Proc. Nat. Acad. Sci. 78, 4078 (1981).

122. Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G. and Stein, J., Proc. Nat. Acad. Sci. 79, 1795 (1982).

123. Heintz, N., Zernik, M. and Roeder, R.G., Cell 24, 661 (1981).

124. Cohen, S.N., Chang, A.C.Y., Boyer, H.W. and Helling, R.G., Proc. Nat. Acad. Sci. 70, 3240 (1973).

125. Morrow, J.F., Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Goodman, H.M. and Helling, R.B., Proc. Nat. Acad. Sci. 71, 1743 (1974).

126. Shortle, D., DiMaio, D. and Nathans, D., Ann. Rev. Genet. 15, 265 (1981).

127. Gurdon, J.B., Birnstiel, M.L. and Speight, V.A., Biochem. Biophys. Acta 174, 614 (1969).

128. Laskey, R.A. and Gurdon, J.B., Eur. J. Biochem. 37, 467 (1973).

129. Mertz, J.E. and Gurdon, J.B., Proc. Nat. Acad. Sci. 74, 1502 (1977).

130. Brown, D.D. and Gurdon, J.B., Proc. Nat. Acad. Sci. 74, 2064 (1977).

131. Colman, A., Eur. J. Biochem. 57, 85 (1975).

132. McKnight, S.L. and Gavis, E.R., Nuc. Acids Res. 8, 5931 (1980).

133. Kressmann, A., Clarkson, S.G., Telford, J.L., and Birnstiel, M.L., Cold Spring Harbor Symp. Quant. Biol. 42, 1077 (1977).

134. McKnight, S.L., Gavis, E.R. and Kingsburg, R., Cell 25, 385 (1981).

135. Bantei, N., Boll, W. and Weissmann, C., Nature 281, 40 (1979).

136. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R., Cell 16, 777 (1979).

137. Wold, B., Wigler, M., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R., Proc. Nat. Acad. Sci. 76, 5684 (1979).

138. Pellicer, A., Robins, D., Wold, B., Sweet, R., Jackson, J., Lowy, I., Roberts, J.M., Sim, G.K., Silverstein, S. and Axel, R., Science 209, 1414 (1980).

139. Hamer, D.H. and Leder, P., Nature 281, 35 (1979).

140. Mulligan, R.C., Howard, B.H. and Berg, P., Nature 277, 108 (1979).

141. Hamer, D.H., Kaehler, M. and Leder, P., Cell 21, 697 (1980).

142. Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T., Cell 27, 279 (1981).

143. Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G., Cell 18, 469 (1979).

144. Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L., Proc. Nat. Acad. Sci. 77, 3855 (1980).

145. Grosschedl, R. and Birnstiel, M.L., Proc. Nat. Acad. Sci. 77, 7102 (1980).

146. Hamer, D.H., Davoli, D., Thomas, C.A. and Fareed, G.C., J. Molec. Biol. 112, 155 (1977).

147. Upcroft, P., Skolnik, H., Upcroft, J.A., Solomon, D., Khoury, G., Hamer, D.H. and Fareed, G.C., Proc. Nat. Acad. Sci. 75, 2117 (1978).

148. Gluzman, Y., Cell 23, 175 (1981).

149. Myers, R.M. and Tjian, R., Proc. Nat. Acad. Sci. 77, 6491 (1980).

150. Weintgartner, B. and Keller, W., Proc. Nat. Acad. Sci 78, 4092, (1981).

151. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corse, P., Kedinger, C. and Chambon, P., Science 209, 1406 (1980).

152. Wasylyk, B., Kedinger, C., Corden, J., Brison, O. and Chambon, P., Nature 285, 367 (1980).

153. Luse, D.S. and Roeder, R.G., Cell 20, 691 (1980).

154. Grosveld, B.C., Shewmaker, C.K., Jat, P. and Flavell, R.A., Cell 25, 215 (1981).

155. Proudfoot, N.J., Shander, M.H.M., Mauley, J.L., Gefter, M.L. and Maniatis, T., Science 209, 1329 (1980).

156. Mueller, C., Graessmann, A. and Graessmann, M., Cell 15, 579 (1978).

157. Ostro, M.J., Giacomoni, D., Lavelle, D., Paxton, W. and Dray, S., Nature 274, 921 (1978).

158. Dimitriadis, G.J., Nature 274, 923 (1978).

159. Loyter, A., Zakai, N. and Kulka, R.G., J. Cell Biol. 66, 292 (1975).

160. Roeder, R.G. in "RNA polymerase", eds. Losick, R. and Chamberlin, M. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.), pp. 285-329 (1976).

161. Jaehning, J.A. and Roeder, R.G., J. Biol. Chem. 252, 8753 (1977).

162. Peterson, R.C., Doering, J.L. and Brown, D.D., Cell 20, 131 (1980).

163. Carrol, D. and Brown, D.D., Cell 7, 477 (1976).

164. Ross Miller, J., Cartwright, E.M., Brownlee, G.G., Fedoroff, N.V. and Brown, D.D., Cell 13, 717 (1978).

165. Korn, L.J. and Brown, D.D., Cell 15, 1145 (1978).

166. Denis, H., Wegnez, M. and Willem, R., Biochimie 54, 1189 (1972).

167. Sakonju, S., Bogenhagen, D.F. and Brown, D.D., Cell 19, 13 (1980).

168. Bogenhagen, D.F., Sakonju, S. and Brown, D.D., Cell 19, 27 (1980).

169. Engelke, D.R., Ng, S.Y., Shastry, B.S. and Roeder, R.G., Cell 19, 717 (1980).

170. Galas, D. and Schmitz, A., Nuc. Acids Res. 5, 3157 (1978).

171. Ng, S.Y., Parker, C.S. and Roeder, R.G., Proc. Nat. Acad. Sci. 76, 136 (1979).

172. Parker, C.S. and Roeder, R.G., Proc. Nat. Acad. Sci. 74, 44 (1977).

173. Sklar, V.E.F. and Roeder, R.G., Cell 10, 405 (1977).

174. Gottesfeld, J. and Bloomer, L.S., Cell 28, 781 (1982).

175. Garber, R.L. and Gage, L.P., Cell 18, 817 (1979).

176. Hagenbuchle, O., Larson, D., Hall, G.I. and Sprague, K.U., Cell 18, 1217 (1979).

177. Sprague, K.U., Larson, D. and Morton, D., Cell 22, 171 (1980).

178. Bogenhagen, D.F. and Brown, D.D., Cell 24, 261 (1981).

179. Rosenberg, M. and Court, D., Ann. Rev. Genet. 13, 319 (1979).

180. Goldberg, M., Ph.D. Thesis, Stanford University, Stanford, California (1979).

181. Efstratiadis, A., Pasakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J., Cell 21, 653 (1960).

182. Grosschedl, R. and Birnstiel, M.L., Proc. Nat. Acad. Sci. 79, 297 (1982).

183. Benoist, C. and Chambon, P., Proc. Nat. Acad. Sci. 77, 3865 (1980).

184. Bendig, M.H., Thomas, T. and Folk, W.R., Cell 20, 401 (1980).

185. Wasylyk, B. and Chambon, P., Nuc. Acids Res. 9, 1813 (1981).

186. Pribnow, D., Proc. Nat. Acad. Sci. 72, 784 (1975).

187. Miller, J.H. in "The Operon", eds. Miller, J.H. and Reznikoff, W.S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) pp. 31-88 (1978).

188. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P., Nuc. Acids Res. 8, 127 (1980).

189. Gruss, P., Dhar, R. and Khoury, G., Proc. Nat. Acad. Sci. 78, 943 (1981).

190. Benoist, C. and Chambon, P., Nature 290, 304 (1981).

191. Sures, I., Levy, S. and Kedes, L.H., Proc. Nat. Acad. Sci. 77, 1265 (1980).

192. Hentschel, C., Irminger, I.C., Bucher, P. and Birnstiel, M.L., Nature 285, 147 (1980).

193. Hagenbuchle, O., Santer, M., Steitz, J.A. and Mans, R.J., Cell 13, 551 (1978).

194. Baralle, F.E. and Brownlee, G.G., Nature 274, 84 (1978).

195. Breathnach, R. and Chambon, P., Ann. Rev. Biochem. 50, 349 (1981).

196. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J., Nature 295, 380 (1982).

197. Proudfoot, N.J. and Brownlee, G.G., Nature 263, 211 (1976).

198. Busslinger, M., Portmann, R. and Birnstiel, M.L. Nuc. Acids Res. 6, 2997 (1979).

199. Proudfoot, N.J., Nature 298, 516 (1982). News & Views.

200. Birchmeier, C., Grosschedl, R. and Birnstiel, M.L., Cell 28, 739 (1982).

201. Harvey, R.P., Robins, A.J. and Wells, J.R.E., Nuc. Acids Res. 10, 7851 (1982).

202. Sierra, F., Leza, A., Marashi, F., Plumb, M., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G.S. and Stein, J.L., Biochem. Biophys. Res. Comm. 104, 785 (1982).

203. Rickles, R., Marashi, F., Sierra, F., Wells, J., Stein, J. and Stein, G., Proc. Nat. Acad. Sci. 79, 749 (1982).

204. Lichtler, A.C., Sierra, F., Clark, S., Wells, J.R.E., Stein, J.L. and Stein, G.S., Nature 298, 195 (1982).

205. Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kieter, D.O., Moore, D.D., Schuman, J.W., Sheldon, E.L. and Smithies, O., Science 196, 161 (1977).

206. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C. and Quon, C, Cell 15, 687 (1978).

207. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T, Cell 15, 1157 (1978).

208. Clewell, D.B. and Helinski, D.R., Biochemistry 9, 4428 (1970).

209. Williams, B. and Blattner, F.J., Virol. 29, 555 (1979).

210. Benton, W.D. and Davis, R.W., Science 196, 180 (1977).

211. McMaster, G.K., Beard, P., Engers, H.D. and Hirt, B., J. Virol. 38 317 (1981).

212. Denhardt, D.T., Biochem. Biophys. Res. Comm. 23, 641 (1966).

213. Droven, J., J. Mol. Biol. 96, 171 (1980).

214. Ross, H.H., Anal. Chemistry 41, 1260 (1969).

215. Southern, E.M., J. Mol. Biol. 98, 503 (1975).

216. Stein, G.S., Park, W.D., Thrall, C.L., Mans, R.J. and Stein, J.L., Nature 257, 764 (1975).

217. Stein, J.L., Thrall, C.L., Park, W.D., Mans, R.J. and Stein, G.S., Science 189, 557 (1975).

218. Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C., and Richardson, C.C., J. Biol. Chem. 243, 4543 (1968).

219. Grunstein, M. and Hogness, D.S., Proc. Nat. Acad. Sci. 72, 3961 (1975).

220. Maxam, A.M. and Gilbert, W., Methods in Enzymol. 65, 499 (1980).

221. McDonnell, M.W., Simon, M.N. and Studier, F.W., J. Mol. Biol. 110, 119 (1977).

222. Rave, N., Crkvenjakov, R. and Boedtker, H., Nuc. Acids Res. 6, 3559 (1979).

223. Tullis, R.H. and Rubin, H., Anal. Biochem. 107, 260 (1980).

224. Wahli, W., Ryffel, G.U., Wyler, T., Jaggi, R.B., Weber, R. and Dawid, I.B., Dev. Biol. 67, 371 (1978).

225. Gray, H.B., Ostrander, D.A., Hodnett, J.L., Legerski, R.J. and Robberson, D.L., Nuc. Acids Res. 2, 1459 (1975).

226. Melli, M., Whitfield, C., Rao, K.V., Richardson, M. and Bishop, J.O., Nature 231, 8 (1971).

227. Wilson, M.C. and Melli, M.J., J. Mol. Biol. 110, 551 (1977).

228. Kunkel, N.S. and Weinberg, E.S., Cell 14, 313 (1978).

229. Denney, R.M. and Yanofsky, C., J. Mol. Biol. 64, 319 (1972).

230. Turner, P.C. and Woodland, H.R., Nuc. Acids Res. 10, 3769 (1982).

231. Davidson, E.H., Galau, G.A., Angerer, R.C. and Britten, R.J., Chromosoma 51, 253 (1975).

232. Houck, C.M., Rinehart, F.P. and Schmid, C.W., J. Mol. Biol. 132, 289 (1979).

233. Tashima, M., Calabretta, B., Torreli, G., Scofield, M., Maizel, A. and Saunders, G.F., Proc. Nat. Acad. Sci. 78, 1508 (1981).

234. Pan, J., Elder, J.T., Duncan, D.H. and Weissman, S.M., Nuc. Acids Res. 9, 1151 (1981).

235. Elder, J.T., Pan, J., Duncan, C.H. and Weissman, S.M., Nuc. Acids Res. 9, 1171 (1981).

236. Calabretta, B., Robberson, D.L., Maizel, A.L. and Saunders, G.F., Proc. Nat. Acad. Sci. 78, 6003 (1981).

237. Ish-Horowicz, D. and Burke, J.F., Nuc. Acids Res. 9, 2989 (1981).

238. Bolivar, F., Rodriquez, R.L., Greene, P., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S., Gene 2, 95 (1977).

239. Razin, S.V., Chernokhvostov, V.V., Roodyn, A.V., Zbarsky, I.B. and Georgiev, G.P., Cell 27, 65 (1981).

240. Detke, S., Stein, J.L. and Stein, G.S., Nuc. Acids Res. 5, 1515 (1978).

241. Sutcliffe, J.G., Nuc. Acids Res. 5, 2721 (1978).

242. Tjian, R., Cell 26, 1 (1981).

243. Banerji, J., Rusconi, R. and Schaffner, W., Cell 27, 299 (1981).

244. Grunstein, M., Diamond, K.E., Knoppel, E. and Grunstein, J.E., Biochemistry 20, 1216 (1981).

245. Stephenson, E.C., Erba, H.P. and Gall, J.G., Nuc. Acids Res. 9, 2281 (1981).

246. Moorman, A.F.M., de Boer, P.A.J., de Laaf, R.T.M., Van Dongen, W.M.A.M. and Destree, O.H.J., FEBS lett. 136, 45, (1981).

247. Fritsch, E.F., Lawn, R.M. and Maniatis, T., Cell 19, 959 (1980).

248. Lauer, J., Shen, C.K.J. and Maniatis, T., Cell 20, 119 (1980).

249. Hereford, L.M., Osley, M.A., Ludwig, J.R. and McLaughlin, C.S. Cell 24, 367 (1981).

250. Wasylik, B., Kedinger, C., Corden, J., Brison, O. and Chambon, P. Nature 285, 367 (1980).

251. Grosschedl, R. and Birnstiel, M.L., Proc. Nat. Acad. Sci. 77, 1432 (1980).

BIOGRAPHICAL SKETCH

Luis Felipe Sierra Tirado was born in Santiago, Chile, on October 4, 1953. He graduated from the Colegio del Verbo Divino in Santiago in 1970. He then entered the school of Ciencias Quimicas (Chemical Sciences) at the Universidad de Chile. After completion of his thesis project entitled "Isolation and Characterization of Ribonucleic Acids Released by Tumor Cells in Culture," under the direction of Dr. Jose Manuel Ojeda, he received the degrees of Bioquimico (Biochemist) and Licenciado en Bioquimica (Licensed in Biochemistry) from the Universidad de Chile in 1977. After working at the virology unit of the School of Medicine of the Universidad de Chile for a year, he entered in 1978 the graduate program at the Department of Biochemistry and Molecular Biology of the University of Florida. He has accepted a post-doctoral position with Dr. Ueli Schibler at the Swiss Institute for Cancer Research, in Lausanne, Switzerland.

Felipe enjoys spending his free time painting, and while in Chile, he participated in several exhibitions, was invited to hold a one-man exhibition, and won several minor art awards. He married Robin Ann Walter in December 1982.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gary S. Stein, Chairman
Professor of Biochemistry and
Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Janet L. Stein, Co-Chairman
Associate Professor of Immunology
and Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard P. Boyce
Professor of Biochemistry and
Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William K. Holloman
Associate Professor of Immunology
and Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Peter M. McGuire
Assistant Professor of Biochemistry
and Molecular Biology

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

April 1983

Dean, College of Medicine 3/30/83

Madelyn Lockhart
Dean for Graduate Studies and
Research